MAIMON WORKING PAPER No. 17 SEPTEMBER 2025

MEASUREMENT IN HTA: PART 1

FROM EPISTEMIC IGNORANCE TO WILLFUL NEGLECT

Paul C Langley Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis MN

ABSTRACT

Health technology assessment (HTA) emerged in the 1970s promising scientific discipline in healthcare resource allocation. Its central device, the quality-adjusted life year (QALY), purported to integrate length and quality of life into a single index. Health economists constructed QALYs by eliciting ordinal preferences through time trade-off and standard gamble techniques, scaling these "utilities" between zero (death) and one (full health), and multiplying them by duration. This appeared to provide a universal currency for outcomes and, when combined with costs, a basis for incremental cost-per-QALY ratios to guide coverage thresholds. Agencies such as NICE, PBAC, and ICER institutionalized the QALY and the accompanying reference case model, presenting HTA as a triumph of rational rationing.

Yet this triumph concealed a foundational error. The utilities underlying QALYs are ordinal rankings, lacking the interval and ratio properties required by measurement theory. Multiplying them by time produced not a measure but an incoherent hybrid. Reference case models compounded the error by embedding these pseudo-numbers in simulations that could never yield falsifiable claims. Judged against representational measurement theory (RMT) and the standards of normal science, HTA's quantitative core collapses.

This first of two papers examines how such a collapse could occur despite the availability of mature standards of measurement. By the 1970s, Stevens' typology of nominal, ordinal, interval, and ratio scales was widely disseminated; Suppes, Luce, Tukey, and their collaborators had codified the axioms of RMT in Foundations of Measurement *(1971); and Rasch had demonstrated how ordinal responses could be transformed into interval measures under strict conditions of unidimensionality and invariance. These were not esoteric advances but widely known landmarks in psychology, statistics, and the social sciences. Stevens' warning that "failure to observe this principle leads to nonsense" was ignored, and HTA pressed forward with precisely such nonsense, presenting ordinal utilities as if they were ratio measures.*

The evidence suggests this was not innocent ignorance but willful neglect. Policy makers wanted a single number to justify rationing decisions, and health economists eager for relevance provided one. The QALY's appeal lay in its rhetorical and institutional utility, not its scientific warrant. It offered decimals, thresholds, and apparent comparability, while sidestepping the categorical barrier of unidimensionality. By design, measurement was displaced by expedience.

Part I documents this genealogy of failure, showing how HTA was founded in defiance of known measurement standards. Part II turns to the survival of the QALY/reference-case complex, examining how relativism, institutional authority, and curricular omission entrenched a memeplex in which political usefulness eclipsed scientific validity.

INTRODUCTION

Health technology assessment (HTA) was established in the 1970s to bring rationality and scientific discipline to healthcare resource allocation. Its promise was that competing therapies could be compared in a common metric, producing evidence that was rigorous, comparable, and actionable [1]. The device chosen to deliver this promise was the quality-adjusted life year (QALY), which purported to integrate both length and quality of life into a single number. Health economists operationalized the QALY by "valuing" health states through preference-elicitation methods such as the time trade-off and the standard gamble. These utilities, conventionally scaled from zero ("dead") to one ("full health"), were then multiplied by duration to yield QALYs.

The attraction was obvious: a single currency of outcomes, apparently permitting comparisons across diseases and interventions. Cost-effectiveness ratios could then be derived, producing incremental cost-per-QALY statistics to rank therapies and set coverage thresholds. With this, HTA appeared to have solved the rationing problem. The QALY and its reference case model became cornerstones of the field, institutionalized by agencies such as NICE.

Yet beneath this apparent triumph lay a fatal error. The utilities at the core of the QALY were not measures but ordinal preference rankings, stripped of the properties required by measurement theory. Multiplying these pseudo-numbers by time produced not a measure but an incoherent hybrid [2]. The reference case model compounded the error by embedding non-measures in simulations that generated ratios with no empirical warrant. Judged against representational measurement theory (RMT) and the standards of normal science, the evolution of objective knowledge, the entire edifice collapses [3 4].

This paper, Part 1 of two papers, asks how this denial of measurement theory could have occurred. By the 1970s the required rules for measurement, both for manifest and latent constructs had been finalized with the axioms of representational measurement theory and the integration of the Rasch measurement model. Yet HTA persevered in a futile exercise to establish QALYs as the Holy Grail of resource allocation. Was it epistemic ignorance, an innocent failure to recognize measurement standards, or willful neglect, a conscious decision to ignore them for policy convenience? The distinction matters. By the time HTA institutionalized the QALY, Stevens' 1946 typology had long clarified the limits of ordinal data, Rasch's 1960 model had shown how to transform ordinal responses into interval scales, and the axioms of RMT had been codified. They were widely known in psychology, statistics, and the social sciences [5]. As Stevens warned, "Failure to observe this principle leads to nonsense." Yet HTA pressed ahead with precisely such nonsense, embedding ordinal utilities in arithmetic that was never defined.

The more plausible reading is that the decision was motivated less by ignorance than by expedience. Policy makers wanted a single metric to justify rationing decisions. Health economists, eager to be relevant, provided one. The QALY's numbers looked scientific: they had decimals, thresholds, and could be plugged into ratios. That they lacked any measurement warrant

was subordinated to their political and institutional usefulness [6]. What began, at best, as epistemic ignorance soon hardened into willful neglect, entrenched by agencies, guidelines, and journals until it became the orthodoxy of HTA.

Part II of this study takes as axiomatic that the QALY was impossible from the start, disqualified by its failure to meet the standard of unidimensionality. To sustain it, advocates had to ignore the requirements of measurement theory and proceed as if valuing health state descriptions were legitimate. From there, HTA descended into what can only be described as relativistic measurement madness. The time trade-off technique, the construction of multiattribute instruments, the derivation of "utilities" through arbitrary algorithms, the multiplication of these pseudo-numbers with time to yield the QALY, and finally the embedding of this chimera in the reference case model; all are examined and deconstructed. Their inevitable failure is laid bare.

STATUS QUO ANTE: THE KNOWN MEASUREMENT STANDARDS

When health technology assessment began its career in the 1960s and 1970s, its architects acted as though no standards existed for what counts as a measure. The historical mythology of the field often presents it as an innovation born of necessity, improvising methods to meet pressing policy demands. But this is a distortion. By that time, the standards of measurement were not only available but well established, widely cited, and intellectually mature. Three strands of work had already converged: Stevens' pragmatic typology of scale types, the axiomatic program of representational measurement theory (RMT), and Rasch's probabilistic model for latent trait measurement. Together they provided a comprehensive framework for determining when numbers legitimately represent empirical attributes. HTA ignored them all.

Stevens and the Nonsense of Misuse

The first strand was Stevens' 1946 article *On the Theory of Scales of Measurement*, published in *Science*. It remains one of the most cited papers in the social sciences. By the time HTA was taking shape, it had been in circulation for over twenty years, absorbed into textbooks and training programs across psychology, education, and the social sciences. Its framework was no esoteric curiosity but a widely recognized standard.

Stevens' typology distinguished nominal, ordinal, interval, and ratio scales according to the permissible transformations each supported. Nominal scales classify, ordinal scales rank, interval scales allow subtraction but lack a true zero, and ratio scales permit all arithmetic because they preserve order, equal intervals, and an absolute zero. Stevens' pragmatic genius was not only to classify but to warn against error:

> "The operations that can legitimately be applied to a given class of numbers depend upon the properties of the scale on which the numbers are based. Failure to observe this principle leads to nonsense." (*Science* 103, 1946: 677–680).

The warning could not have been clearer. If numbers assigned to preferences or states are ordinal, then arithmetic operations such as averaging, subtraction, or multiplication are illegitimate. To

treat them as though they sustain those operations is to generate nonsense disguised as measurement.

HTA's central constructs were in open defiance of this principle. Utilities derived from time trade-off and standard gamble are ordinal rankings of preference. To multiply these by time to create QALYs is exactly the kind of error Stevens had condemned. By the time health economists were formalizing utilities in the 1970s, Stevens' dictum had been recognized for two decades. They could not plausibly claim ignorance.

From Stevens to the Axiomatic Program: Suppes, Luce and Tukey

Stevens' typology was never intended as the final word on measurement. It was a pragmatic framework, meant to caution scientists against illegitimate arithmetic. But almost immediately, a more rigorous program began to develop that sought to place measurement on firm logical and mathematical foundations. The most important early contributor was Patrick Suppes.

In the 1950s and 1960s, Suppes published a series of papers that sought to formalize what it meant for a numerical assignment to represent an empirical relational structure. His 1951 work *A Set of Independent Axioms for Extensive Quantities* was foundational, offering axioms for extensive measurement that made clear how additivity could be justified [7]. By 1971, *Foundations of Measurement* consolidated these advances. Suppes argued that measurement was not a matter of convention or convenience but a matter of preserving empirical structure. If empirical comparisons could not support operations like addition, then no numerical representation could make them legitimate.

The importance of Suppes' contribution lies in his insistence on axiomatic clarity. Where Stevens had warned pragmatically that misuse leads to nonsense, Suppes showed precisely why: because arithmetic has meaning only when the empirical relations themselves admit such structure. This was a decisive step toward closing the gap between qualitative observation and quantitative representation.

The next milestone came with Duncan Luce and John Tukey's 1964 paper *Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement* [8]. Here the concern was not with single extensive quantities, like length or weight, but with attributes that could be combined. Luce and Tukey showed that under certain axioms, most centrally the cancellation axioms, it is possible to assign numbers to pairs of attributes in such a way that the ordering of empirical combinations is preserved by numerical addition.

This was an extraordinary advance. It demonstrated that measurement could extend beyond the obvious physical quantities to complex relational structures, provided that the empirical data satisfied strict conditions. But it also underscored how demanding those conditions were. Cancellation, in particular, requires consistency across trade-offs: if A combined with B outranks C combined with D, then comparable decompositions must preserve that ordering. Without such consistency, additivity is not justified.

For HTA, this was fatal. Health state descriptions—mobility combined with pain combined with psychological distress—do not satisfy cancellation. Preferences expressed over such profiles do not display the consistency required for conjoint measurement. The assumption that they could be treated as additive indices was contradicted by the very axioms Luce and Tukey had articulated.

By the time the first volume of *Foundations of Measurement* appeared in 1971, these insights had been brought together. The work of Krantz, Luce, Suppes, and Tversky codified the entire representational program, setting down the conditions under which measurement is possible, the structure-preserving mappings that define it, and the theorems that guarantee uniqueness. The appearance of this volume marked the culmination of decades of work. The standards for legitimate measurement were now explicit and beyond dispute.

For HTA, the implications were devastating. The decision to "value" health states through preference exercises such as time trade-off or standard gamble ignored this entire trajectory. By the early 1970s, it was no longer possible to claim ignorance of Stevens' warning, Suppes' axioms, or Luce and Tukey's demonstration of conjoint measurement. The standards of measurement had been recognized for decades and were codified in rigorous form. By 1971, with *Foundations of Measurement*, the case was closed. HTA wasn't improvising in a vacuum but defying settled standards. Yet HTA pressed ahead, treating ordinal preferences as though they were interval or ratio measures. The genealogy of the QALYs was, from its inception, a defiance of the known standards of science.

UNIDIMENSIONALITY AS THE FATAL CONSTRAINT

Among the axioms of representational measurement, the demand for unidimensionality is especially fatal for HTA's project. A valid measure must capture variation in a single attribute along a continuum. Interval and ratio properties are meaningful only within such a unidimensional space. Once multiple attributes are collapsed into a composite index, the logic of measurement collapses with it.

This requirement had been clear since Stevens' landmark 1946 paper *On the Theory of Scales of Measurement*. Stevens distinguished ordinal scales, which provide only rank order, from interval and ratio scales, which add continuity. As he emphasized:

> "Interval scales not only permit the rank-ordering of the objects of study but also the determination of the equality of the intervals or differences between them. Ratio scales possess all the properties of interval scales and in addition they identify or define the absolute zero of the scale." (*Science* 103, 1946: 678).

Continuity was thus the defining property that elevated a ranking into a measure. Equal intervals presuppose a unidimensional continuum: without it, arithmetic cannot be justified. Ratio scales add further power by anchoring the continuum at a true zero, permitting proportional comparisons such as "twice as long." Stevens' framework left no ambiguity—ordinal rankings cannot be treated as though they had interval or ratio properties. To do so was, in his words, to generate "nonsense."

Unidimensionality is what secures this continuity. For an interval or ratio scale to be meaningful, the numbers must represent differences along a single attribute. Length is measurable because all rods lie along the dimension of extension; time is measurable because all events occupy the same temporal line. Multiattribute composites, by contrast, have no such foundation. Collapsing pain, mobility, anxiety, and social functioning into one number produces not a continuum but a contrivance. No test can demonstrate additivity across such disparate domains, because no unidimensional structure exists.

This is why Rasch's model (see below) was so crucial. It provided the first rigorous method for extracting a unidimensional latent trait from ordinal responses. Items that do not conform to the underlying construct are discarded, preserving unidimensionality. The logit transformation then yields an interval continuum in which equal distances correspond to equal differences in the trait. By insisting on unidimensionality, Rasch operationalized what Stevens had already signaled: continuity is the dividing line between rank and measure.

For HTA, the implications are devastating. Utilities built on multiattribute indices such as the EQ-5D can never satisfy unidimensionality. The attributes they collapse are not aligned along a single continuum. Their weights are conventional, not empirical. They yield at best ordinal rankings and at worst incoherent numbers. To treat them as interval or ratio measures is to ignore the very distinction Stevens insisted upon and Rasch later demonstrated in practice.

This is why only two kinds of measures are admissible in HTA. The first are linear ratio scales for manifest attributes—time, costs, hospital days, and resource units—quantities that vary on an obvious continuum with a true zero. The second are Rasch logit ratio scales for latent traits such as need fulfillment or quality of life, extracted from ordinal responses under Rasch constraints. Both satisfy unidimensionality, and both sustain continuity. Everything else collapses at the threshold.

The implication is stark. Utilities and the QALY were disqualified at birth. They ignore unidimensionality, deny continuity, and yet claim to be measures. Once continuity is taken seriously, only linear ratios and Rasch-based ratios remain. Here lies the fatal constraint: unidimensionality was not a technical refinement but a categorical boundary condition. Cross it, and you are no longer doing measurement; only numerical storytelling.

RASCH AND THE TRANSFORMATION OF ORDINAL RESPONSES

If multiattribute indices are excluded, how can subjective constructs be measured? Rasch provided the answer. In 1960 Georg Rasch published *Probabilistic Models for Some Intelligence and Attainment Tests*, introducing a model that would revolutionize measurement in the human sciences. His aim was not to accommodate data within a flexible statistical framework but to construct a model that could sustain the same rigor as measurement in the physical sciences. Rasch was not motivated by health economics or policy demands but by a deeper concern: how to establish measurement in psychology and education where observations are inherently probabilistic.

Rasch's model was developed independently of representational measurement theory. It did not arise from the axiomatic program of Suppes, Luce, Tukey, or Krantz, though it would later be recognized as uniquely consistent with their work. Instead, Rasch was addressing a practical problem: test data consist of right or wrong answers, or more generally ordered categorical responses, and these are subject to chance variation. Classical test theory treated scores as though they were interval measures, but this assumption lacked justification. Rasch sought a model that would allow the transformation of these ordinal observations into a scale with the properties of measurement.

The elegance of Rasch's solution lay in its simplicity. He proposed that the probability of a correct response could be expressed as a logistic function of the difference between two parameters: the ability of the person and the difficulty of the item. If a person's ability equals an item's difficulty, the probability of success is 0.5; if ability exceeds difficulty, the probability rises above 0.5; if it falls short, the probability drops below 0.5. This formulation placed persons and items on the same continuum. It implied that both could be estimated independently and located on a shared scale of measurement.

What made this model groundbreaking was not simply that it fit data, *but that it demanded the data fit the model*. Rasch was explicit that his approach was not statistical curve-fitting but measurement. The Rasch model defines the conditions under which observations can be transformed into interval-level measures. Its core properties demonstrate this rigor:

- Unidimensionality: the model assumes that responses reflect variation in a single latent trait. This ensures that the resulting scale has a coherent meaning; all items measure the same underlying construct.
- Invariance: item difficulties are estimated independently of the sample of persons, and person abilities are estimated independently of the set of items. This property mirrors physical measurement, where the length of a ruler does not depend on which objects are measured, nor does the size of an object depend on which ruler is used.
- Interval scaling: by applying the logit (log-odds) transformation, the Rasch model converts ordinal probabilities into an interval continuum. Equal differences in logits correspond to equal differences in the latent trait, fulfilling a core requirement of measurement.

By the 1970s, extensions of the Rasch model to polytomous data, such as the rating scale model and the partial credit model, made it possible to apply the same principles to Likert-type items and ordered categories [9] The significance was profound: Rasch provided the practical demonstration of how subjective, ordinal responses could be transformed into interval measures that met the axioms of representational measurement. Rasch, it must be emphasized, isn't simply "one possible model" but the *necessary* one for latent constructs

This achievement marked Rasch's unique status in the history of measurement. Other statistical models, such as factor analysis or item response theory in its 2- and 3-parameter forms, allowed parameters to vary freely to maximize fit. Rasch rejected this flexibility. His model is deliberately restrictive because only under those restrictions do the axioms of measurement hold. Where conventional models aim to describe data, Rasch aims to establish measurement. This distinction is what Wright would later emphasize when he declared that Rasch is the unique model consistent

with representational measurement theory. Its probabilistic logic naturally yields the additive structure that the axioms demand.

The independence of Rasch's development is itself striking. Without reference to Stevens' typology or to the axiomatic program of RMT, Rasch nonetheless converged on a solution that fulfilled their requirements. He demonstrated that it was possible to move from ordinal data to interval measurement, but only when the strict assumptions of the Rasch model are satisfied. In doing so, he bridged the gap between the philosophy of measurement and the practice of social science.

For HTA, the lesson could not be clearer. If subjective attributes such as quality of life are to be measured, they must be modeled within a framework that guarantees unidimensionality, invariance, and interval scaling. Multiattribute composites like the EQ-5D fail all of these tests. Rasch shows that measurement is possible, but only when it respects the axioms. By the time QALYs were being formalized, Rasch's work had already been extended and was widely known in psychometrics and educational measurement. To ignore it was not simply to miss a technical refinement; it was to turn away from the only available method for making subjective constructs measurable.

Rasch's contribution was thus both conceptual and practical. Conceptually, he demonstrated that measurement in the social sciences is not impossible; it requires strict conditions. Practically, he provided the tools to implement those conditions. That HTA never engaged with Rasch is therefore not an oversight but an indictment. It reveals a field committed to policy expedience rather than to the standards of measurement science.

RASCH AND WRIGHT: THE UNIQUE MEASUREMENT MODEL

The decisive link between Rasch and RMT was made explicit by. Wright in 1977 [10]. Wright argued that the Rasch model is not simply one among many statistical models but the unique framework that fulfills the axioms of representational measurement. Its probabilistic form naturally yields the additive structure demanded by conjoint measurement.

As Wright put it:

> "The purpose of this section is to make clear that Rasch's model fulfills the axioms of representational measurement, not because it was crafted in deference to them, but because its probabilistic logic naturally yields the structure those axioms describe." (*Solving Measurement Problems with the Rasch Model*, 1977).

This was a watershed statement. It meant that Rasch was not a convenient approximation but the only model consistent with measurement theory. It alone provided the bridge from ordinal observation to interval measurement in the social sciences. By 1977, no one could plausibly claim ignorance of the standards.

Against this background, the decision to "value" health states using multiattribute indices looks indefensible. Stevens had warned against treating ordinal numbers as interval. RMT had set down

the axioms showing why multiattribute composites could not be measures. Rasch had provided the method for constructing legitimate unidimensional instruments. Wright had explicitly linked Rasch to the axioms, demonstrating its uniqueness. Yet HTA persevered into what might be described as a relativist yet defeatist measurement universe that denied the RMT axioms for both manifest constructs and latent manifest traits. .

In other words, the *status quo ante* was one in which the requirements of measurement were not just available but mature, explicit, and actionable. To proceed with QALYs in defiance of these standards cannot be excused as epistemic ignorance. By the mid-1970s, when QALYs were being formalized, the field already knew what measurement required. To ignore unidimensionality, to bypass RMT, and to neglect Rasch was willful.

This was a watershed. It demonstrated that Rasch provided the only bridge from ordinal responses to legitimate measurement. By the late 1970s, no one in HTA could plausibly claim ignorance of what measurement required.

MEASUREMENT IS IRRELEVANT BY DESIGN

Against this background, the decision to "value" health states with multiattribute composites looks indefensible. Stevens had warned against treating ordinal data as interval. Suppes had clarified the axioms of extensive measurement. Luce and Tukey had defined the cancellation requirements for conjoint measurement. *Foundations of Measurement* (1971) had codified the axioms. Rasch had provided the practical method. Wright had established Rasch's unique status. The standards were not only available but explicit.

Yet the genealogy of HTA shows that these standards were not just overlooked but deliberately displaced. MacKillop and Sheard's *Quantifying Life: Understanding the history of Quality Adjusted Life Years (QALY)* (2018) confirms this. Their history traces the QALY's rise in the UK and US, highlighting its appeal to policymakers who were searching for a common currency to compare disparate medical interventions. The QALY's power lay in its simplicity: it promised to combine length and quality of life into a single number, producing an apparently objective basis for resource allocation. The attraction was rhetorical and institutional, not methodological. Crucially, their account contains no mention of Stevens, RMT, or Rasch; not because these were obscure, but because they were irrelevant to the project as framed. Measurement was never a concern.

That silence is itself revealing. The absence of discussion of measurement is not a neutral omission; it reflects the priorities of the actors who shaped the field. The imperative was political: to provide governments and health systems with an index that could be used to justify rationing decisions in an era of escalating costs. The ambition was not to build a science of outcomes but to secure administrative traction. In that environment, precision, dimensional homogeneity, or axiomatic coherence were not merely overlooked; they were irrelevant by design. To acknowledge them would have undermined the very possibility of a single index. An absence that continues to this day.

The result was that HTA was born in a space where policy convenience and institutional authority were privileged over scientific validity. By design, the QALY operated as a tool of persuasion rather than measurement. It was a political technology masquerading as a scientific metric. The failure to cite Stevens, Suppes, Luce, Tukey, or Rasch is therefore not incidental. It is evidence that the genealogy of HTA was shaped by willful neglect of science in favor of utility to decision-makers. Measurement was absent not because the standards were unknown, but because their recognition would have rendered the enterprise impossible.

That silence is revealing. It shows that the actors who built HTA never intended to ground their constructs in measurement theory. Their priorities were institutional traction, comparability, and rhetorical force. The fact that the numbers lacked the properties of measures was simply not part of the discourse. The history of HTA is, in this sense, a history of measurement absent by design.

CONCLUSION

Part I has argued that measurement failure in HTA was not an accident of history but a choice. By the time utilities and the QALY were institutionalized, the conditions for legitimate measurement were settled. Stevens had already drawn the boundary between rank and measure; Suppes, Luce, and Tukey had shown why additivity requires cancellation; *Foundations of Measurement* had codified representation and uniqueness; Rasch had given a workable method for transforming ordinal responses into interval scales; and Wright had made explicit Rasch's unique consistency with the axioms. Against this background, the elevation of multiattribute health state valuations to "utilities," and their multiplication by time to produce QALYs, cannot be excused as epistemic innocence. It was willful neglect; a retreat from science.

Once this neglect is recognized, the subsequent architecture of HTA is no longer puzzling. The reference case model formalized the initial error, embedding non-measures in simulations that cannot be falsified and yielding ratios that lack dimensional meaning. What appeared as a unifying metric for rational rationing was, in fact, a political technology: a persuasive index that could be tabulated, thresholds proposed and applied, while remaining untethered to the formal axioms of measurement. The enduring appeal of this technology lay not in its scientific warrant but in its administrative convenience and rhetorical force. Understanding formal measurement was not required; indeed, it would be a distraction. A position that continues to this day.

The remedy is not repair but replacement. Utilities and QALYs cannot be rehabilitated because they never met the preconditions of measurement. If HTA is to operate as science rather than numerical storytelling, its claims must be grounded in the only admissible forms of quantification: linear ratio measures for manifest attributes such as time, costs, and resource use, and Rasch-based logit ratio measures for latent constructs developed under strict unidimensionality and invariance. Claims stated on these foundations can be credible, evaluable, replicable, and falsifiable.

Part II will explain why, despite the availability of these standards, the QALY/reference-case complex not only survived but dominated. The answer lies in the shift from willful neglect to relativism: a memeplex sustained by consensus, authority, and curricular omission, in which truth is displaced by usefulness and dissent is neutralized by procedure. If Part I documents the genesis of measurement failure, Part II will explain its survival under relativism: how consensus, authority,

and curricular omission sustained a memeplex in which truth was displaced by usefulness and dissent neutralized by procedure.

ACKNOWEDGEMENT

Portions of this paper were drafted and edited with assistance from ChatGPT (version 5; OpenAI). The author reviewed, verified, and refined all AI-assisted text and assumes full responsibility for the accuracy, integrity, and originality of the final content.

REFERENCES

[1] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed.) New York: Oxford University Press, 2015

[2] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

[3] Krantz D, Luce R, Suppes P, Tversky A**.** *Foundations of Measurement*, Volumes I–III. New York: Academic Press, 1971 (Vol. I); 1989 (Vol. II); 1990 (Vol. III).

[4] Popper K. *Objective Knowledge: An Evolutionary Approach*. Revised edition. Oxford: Clarendon Press, 1979

[5] Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960. (Expanded ed., Chicago: University of Chicago Press, 1980)

[6] MacKillop E, Sheard S Quantifying Life: Understanding the History of Quality Adjusted Life Years (QALYs).” *Social Science & Medicine*. 2018; 211: 359–366 [see also MacKillop E and Sheard S. *Quantifying Life: Understanding the History of Quality-Adjusted Life Years (QALYs).* Manchester: Manchester University Press, 2018]

[7] Suppes P. A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica* 1951; 10: 163–172.

[8] Luce R, Tukey J. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. J *Math Psychol.* 1964*;* 1(1): 1–27

[9] Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed). New York: Routledge, 2021

[10] Wright B. “Solving Measurement Problems with the Rasch Model.” *Journal of Educational Measurement* 14, no. 2 (1977): 97–116