MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

ISPOR GLOBAL NETWORK: THE ENDORSEMENT OF MEASUREMENT FAILURE - AFRICA NETWORK

Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 4016 FEBRUARY 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF MEASUREMENT INVERSION FAILURE

ABSTRACT

Health technology assessment (HTA) across Africa is expanding through national agencies, university programs, health-economics research groups, professional networks and ISPOR regional chapters. These developments increasingly incorporate economic evaluation, utilities, QALYs, cost-effectiveness modelling and related health economics and outcomes research (HEOR) methods. The present study examines whether the ISPOR-related Africa knowledge base first satisfies the requirements for lawful measurement or instead reproduces measurement inversion, in which arithmetic precedes the establishment of measurement.

The Africa knowledge base was defined as the publicly accessible ISPOR-related regional environment encompassing Algeria, Egypt, Ethiopia, Ghana, Kenya, Morocco, Nigeria, South Africa and Uganda, together with associated academic, governmental, HTA, reimbursement and HEOR activities. A fixed set of 24 canonical statements derived from representational measurement and Rasch measurement principles was applied. Each statement was assigned a categorical endorsement probability and corresponding normalized logit to determine whether the regional knowledge base reinforces the requirements for lawful measurement or the numerical conventions of contemporary HTA.

The interrogation identified a pronounced measurement-inversion profile. True propositions concerning measurement preceding arithmetic, unidimensionality, ratio measurement as a prerequisite for multiplication and the axioms of representational measurement received weak endorsement. In contrast, false propositions supporting conventional HEOR practices including the ratio status and dimensional homogeneity of QALYs, aggregation of QALYs, interval interpretation of preference algorithms and ratio interpretation of summed subjective scores received strong endorsement. Rasch-based propositions concerning latent measurement and possession of attributes received particularly weak endorsement. The resulting pattern indicates that arithmetic, statistical estimation, modelling and preference-based analysis are accepted without first establishing the measurement properties required to justify those operations.

The ISPOR Africa knowledge base reproduces the measurement-failure signature identified in the parent ISPOR framework and other regional HTA environments. The findings should not be interpreted as an African invention of measurement failure, but as evidence of successful adoption of an inherited methodological paradigm in which lawful measurement is not the prior gatekeeper. This is especially consequential as African countries expand HTA capacity. The required transition is to restore the scientific sequence: define the attribute, establish the appropriate lawful measure, permit only admissible arithmetic, and formulate prospectively evaluable and falsifiable

therapy-impact claims. The Maimon Research nine-unit transformation program provides a structured pathway for this transition.

INTRODUCTION: ISPOR GLOBAL NETWORK AFRICA

The ISPOR Global Network provides an international structure through which the Society promotes health economics and outcomes research (HEOR), pharmacoeconomics, health technology assessment (HTA), economic evaluation and their application to healthcare decision-making. Regional networks and country chapters provide an important mechanism for extending the ISPOR methodological framework into local academic, professional and policy environments.

ISPOR formally identifies an Africa Network within its Global Networks structure, while its Regional Chapter Directory groups country chapters geographically under Africa. For the purposes of the present review, the Africa knowledge base is defined by this regional structure and comprises Algeria, Egypt, Ethiopia, Ghana, Kenya, Morocco, Nigeria, South Africa and Uganda**.** Morocco is also identified by ISPOR within its Middle East regional grouping and therefore represents an overlap between the Africa and Middle East knowledge bases.

These country chapters provide a defined regional structure through which ISPOR principles, educational activities and HEOR methodologies can be disseminated and applied. Their importance extends beyond professional networking. As HTA and formal economic evaluation develop across Africa, the methodological framework promoted through ISPOR can influence research, professional education, reimbursement and healthcare policy. The Africa Network therefore provides an important regional knowledge base for assessing whether contemporary HEOR recognizes the requirements of lawful measurement before undertaking quantitative evaluations of therapy impact.

The objective of this study is to determine whether the ISPOR Africa knowledge base endorses the requirements for lawful measurement in the quantitative evaluation of therapy impact or, alternatively, endorses measurement inversion. Using 24 diagnostic statements, the interrogation assesses recognition of representational measurement, scale properties, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, Rasch measurement, utilities, QALYs and falsifiability. The fundamental test is whether measurement is required to precede arithmetic.

The findings demonstrate a consistent endorsement of measurement inversion and, therefore, measurement failure. Foundational requirements for lawful measurement receive low endorsement, including measurement preceding arithmetic, ratio measurement for multiplication and the axioms of representational measurement. In contrast, false statements supporting conventional HEOR practices, including the ratio status, dimensional homogeneity and aggregation of QALYs, receive strong endorsement. The Africa knowledge base therefore reproduces the measurement-failure pattern previously identified within the parent ISPOR framework.

The starting point is simple and inescapable: *measurement precedes arithmetic*. This principle is not a methodological preference but a logical necessity. One cannot multiply what one has not measured, cannot sum what has no dimensional homogeneity, cannot compare ratios when no ratio scale exists. When HTA multiplies time by utilities to generate QALYs, it is performing arithmetic with numbers that cannot support the operation. When HTA divides cost by QALYs, it is constructing a ratio from quantities that have no ratio properties. When HTA aggregates QALYs across individuals or conditions, it is combining values that do not share a common scale. These practices are not merely suboptimal; they are mathematically impossible.

The modern articulation of this principle can be traced to Stevens' seminal 1946 paper, which introduced the typology of nominal, ordinal, interval, and ratio scales [1]. Stevens made explicit what physicists, engineers, and psychologists already understood: different kinds of numbers permit different kinds of arithmetic. Ordinal scales allow ranking but not addition; interval scales permit addition and subtraction but not multiplication; ratio scales alone support multiplication, division, and the construction of meaningful ratios. Utilities derived from multiattribute preference exercises, such as EQ-5D or HUI, are ordinal preference scores; they do not satisfy the axioms of interval measurement, much less ratio measurement. Yet HTA has, for forty years, treated these utilities as if they were ratio quantities, multiplying them by time to create QALYs and inserting them into models without the slightest recognition that scale properties matter. Stevens' paper should have blocked the development of QALYs and cost-utility analysis entirely. Instead, it was ignored.

The foundational theory that establishes *when* and *whether* a set of numbers can be interpreted as measurements came with the publication of Krantz, Luce, Suppes, and Tversky's *Foundations of Measurement* (1971) [2] . Representational Measurement Theory (RMT) formalized the axioms under which empirical attributes can be mapped to numbers in a way that preserves structure. Measurement, in this framework, is not an act of assigning numbers for convenience, it is the discovery of a lawful relationship between empirical relations and numerical relations. The axioms of additive conjoint measurement, homogeneity, order, and invariance specify exactly when interval scales exist. RMT demonstrated once and for all that measurement is not optional and not a matter of taste: either the axioms hold and measurement is possible, or the axioms fail and measurement is impossible. Every major construct in HTA, utilities, QALYs, DALYs, ICERs, incremental ratios, preference weights, health-state indices, fails these axioms. They lack unidimensionality; they violate independence; they depend on aggregation of heterogeneous attributes; they collapse under the requirements of additive conjoint measurement. Yet HTA proceeded, decade after decade, without any engagement with these axioms, as if the field had collectively decided that measurement theory applied everywhere except in the evaluation of therapies.

Whereas representational measurement theory articulates the axioms for interval measurement, Georg Rasch's 1960 model provides the only scientific method for transforming ordered categorical responses into interval measures for latent traits [3]. Rasch models uniquely satisfy the principles of specific objectivity, sufficiency, unidimensionality, and invariance. For any construct such as pain, fatigue, depression, mobility, or need, Rasch analysis is the only legitimate means of producing an interval scale from ordinal item responses. Rasch measurement is not an alternative to RMT; it is its operational instantiation. The equivalence of Rasch's axioms and the axioms of representational measurement was demonstrated by Wright, Andrich and others as early as the 1970s. In the latent-trait domain, the very domain where HTA claims to operate; Rasch is the only game in town [4].

Yet Rasch is effectively absent from all HTA guidelines, including NICE, PBAC, CADTH, ICER, SMC, and PHARMAC. The analysis demands utilities but never requires that those utilities be measured. They rely on multiattribute ordinal classifications but never understand that those constructs be calibrated on interval or ratio scales. They mandate cost-utility analysis but never justify the arithmetic. They demand modelled QALYs but never interrogate their dimensional properties. These guidelines do not misunderstand Rasch; they do not know it exists. The axioms that define measurement and the model that makes latent trait measurement possible are invisible to the authors of global HTA rules. The field has evolved without the science that measurement demands.

How did HTA and then ISPOR miss the bus so thoroughly? The answer lies in its historical origins. In the late 1970s and early 1980s, HTA emerged not from measurement science but from welfare economics, decision theory, and administrative pressure to control drug budgets. Its core concern was *valuing health states*, not *measuring health*. This move, quiet, subtle, but devastating, shifted the field away from the scientific question "What is the empirical structure of the construct we intend to measure?" and toward the administrative question "How do we elicit a preference weight that we can multiply by time?" The preference-elicitation projects of that era (SG, TTO, VAS) were rationalized as measurement techniques, but they never satisfied measurement axioms. Ordinal preferences were dressed up as quasi-cardinal indices; valuation tasks were misinterpreted as psychometrics; analyst convenience replaced measurement theory. The HTA community built an entire belief system around the illusion that valuing health is equivalent to measuring health. It is not.

The endurance of this belief system, forty years strong and globally uniform, is not evidence of validity but evidence of institutionalized error. HTA has operated under conditions of what can only be described as *structural epistemic closure*: a system that has never questioned its constructs because it never learned the language required to ask the questions. Representational measurement theory is not taught in graduate HTA programs; Rasch modelling is not part of guideline development; dimensional analysis is not part of methodological review. The field has been insulated from correction because its conceptual foundations were never laid. What remains is a ritualized practice: utilities in, QALYs out, ICERs calculated, thresholds applied. The arithmetic continues because everyone assumes someone else validated the numbers.

This Logit Working Paper series exposes, through probabilistic and logit-based interrogations of AI large language national knowledge bases, the scale of this failure. The results display a global pattern: true statements reflecting the axioms of measurement receive weak endorsement; false statements reflecting the HTA belief system receive moderate or strong reinforcement. This is not disagreement. It is non-possession. It shows that HTA, worldwide, has developed as a quantitative discipline without quantitative foundations; a confused exercise in numerical storytelling.

The conclusion is unavoidable: HTA does not need incremental reform; it needs a scientific revolution. Measurement must precede arithmetic. Representational axioms must precede valuation rituals. Rasch measurement must replace ordinal summation and utility algorithms. Value claims must be falsifiable, protocol-driven, and measurable; rather than simulated, aggregated, and numerically embellished.

The global system of non-measurement is now visible. The task ahead is to replace it with science.

Paul C Langley, Ph.D

Email: langleylapaloma@gmail.com

DISCLAIMER

This analysis is generated through the structured interrogation of a large language model (LLM) applied to a defined documentary corpus and is intended solely to characterize patterns within an aggregated knowledge environment. It does identify, assess, or attribute beliefs, intentions, competencies, or actions to any named individual, faculty member, student, administrator, institution, or organization. The results do not constitute factual findings about specific persons or programs, nor should they be interpreted as claims regarding professional conduct, educational quality, or compliance with regulatory or accreditation standards. All probabilities and logit values reflect model-based inferences about the presence or absence of concepts within a bounded textual ecosystem, not judgments about real-world actors. The analysis is exploratory, interpretive, and methodological in nature, offered for scholarly discussion of epistemic structures rather than evaluative or legal purposes. Any resemblance to particular institutions or practices is contextual and non-attributive, and no adverse implication should be inferred.

1. INTERROGATING THE LARGE LANGUAGE MODEL

A large language model (LLM) is an artificial intelligence system designed to understand, generate, and manipulate human language by learning patterns from vast amounts of text data. Built on deep neural network architectures, most commonly transformers, LLMs analyze relationships between words, sentences, and concepts to produce contextually relevant responses. During training, the model processes billions of examples, enabling it to learn grammar, facts, reasoning patterns, and even subtle linguistic nuances. Once trained, an LLM can perform a wide range of tasks: answering questions, summarizing documents, generating creative writing, translating languages, assisting with coding, and more. Although LLMs do not possess consciousness or true understanding, they simulate comprehension by predicting the most likely continuation of text based on learned patterns. Their capabilities make them powerful tools for communication, research, automation, and decision support, but they also require careful oversight to ensure accuracy, fairness, privacy, and responsible use

In this Logit Working Paper, "interrogation" refers not to discovering what an LLM *believes,* it has no beliefs, but to probing the content of the *corpus-defined knowledge space* we choose to analyze. This knowledge base is enhanced if it is backed by accumulated memory from the user. In this case the interrogation relies also on 12 months of HTA memory from continued application of the system to evaluate HTA experience. The corpus is defined before interrogation: it may consist of a journal (e.g., *Value in Health*), a national HTA body, a specific methodological framework, or a collection of policy documents. Once the boundaries of that corpus are established, the LLM is used to estimate the conceptual footprint within it. This approach allows us to determine which principles are articulated, neglected, misunderstood, or systematically reinforced.

In this HTA assessment, the objective is precise: to determine the extent to which a given HTA knowledge base or corpus, global, national, institutional, or journal-specific, recognizes and reinforces the foundational principles of representational measurement theory (RMT). The core principle under investigation is that measurement precedes arithmetic; no construct may be treated as a number or subjected to mathematical operations unless the axioms of measurement are satisfied. These axioms include unidimensionality, scale-type distinctions, invariance, additivity, and the requirement that ordinal responses cannot lawfully be transformed into interval or ratio quantities except under Rasch measurement rules.

The HTA knowledge space is defined pragmatically and operationally. For each jurisdiction, organization, or journal, the corpus consists of:

- published HTA guidelines
- agency decision frameworks
- cost-effectiveness reference cases
- academic journals and textbooks associated with HTA
- modelling templates, technical reports, and task-force recommendations
- teaching materials, methodological articles, and institutional white papers

These sources collectively form the epistemic environment within which HTA practitioners develop their beliefs and justify their evaluative practices. The boundary of interrogation is thus not the whole of medicine, economics, or public policy, but the specific textual ecosystem that sustains HTA reasoning. . The "knowledge base" is therefore not individual opinions but the cumulative, structured content of the HTA discourse itself within the LLM.

THE AFRICA HTA KNOWLEDGE BASE

The ISPOR Africa knowledge base is not confined to material formally published by ISPOR or to the activities of individual country chapters. It is better understood as the interconnected body of knowledge through which contemporary health economics and outcomes research (HEOR), pharmacoeconomics and health technology assessment (HTA) are taught, promoted, published and applied across the region. ISPOR provides the international methodological framework, while national chapters, universities, research groups, government agencies, healthcare organizations and individual investigators provide the institutional settings through which that framework is interpreted and implemented.

The knowledge base therefore includes ISPOR regional and country activities; conferences, workshops and educational programs; publications by chapter members and associated researchers; university teaching and research in health economics, pharmacoeconomics and HTA; national HTA initiatives and methodological guidance; reimbursement and formulary assessment frameworks; and published economic evaluations undertaken within the constituent African countries. Particular importance attaches to material addressing cost-effectiveness analysis, utilities, QALYs, preference-based instruments, mapping, patient-reported outcomes, decision-analytic modelling, sensitivity analysis and value assessment. Collectively, these sources reveal the quantitative assumptions that characterize the regional HEOR framework.

This broader definition is essential because knowledge transmission does not occur solely through formal methodological guidance. An ISPOR chapter may publish relatively little itself while its members teach pharmacoeconomics, conduct cost-effectiveness studies, participate in HTA and reimbursement committees, contribute to the development of national guidance and train subsequent generations of practitioners. The relevant knowledge base is therefore the publicly accessible intellectual environment associated with the ISPOR Africa Network rather than a narrowly defined collection of ISPOR documents.

For interrogation purposes, the critical question is what this knowledge base indicates about the status of measurement. Evidence is sought for recognition that quantitative evaluation begins by defining the attribute to be measured and establishing the properties of the resulting scale before arithmetic operations are undertaken. This includes recognition of unidimensionality, interval and ratio scale properties, admissible transformations, dimensional homogeneity, manifest and latent attributes, the requirements for ratio measurement and the role of Rasch measurement for latent attributes. These requirements are contrasted with the treatment of utilities, QALYs, composite scores, mapping, aggregation and simulation modelling as quantitative inputs and outputs.

The interrogation does not assume that every institution or practitioner within Africa holds identical methodological beliefs. Nor does the existence of an ISPOR chapter imply that every publication from that country represents an official ISPOR position. The purpose is instead to characterize the dominant publicly accessible regional knowledge base associated with the ISPOR Africa Network and determine which propositions it would be expected to endorse.

The resulting assessment therefore asks whether the Africa knowledge base reflects a framework in which measurement precedes arithmetic, or one in which numerical operations are accepted without first establishing lawful measurement. The latter constitutes measurement inversion. Where measurement inversion supports quantitative claims of therapy impact, its consequence is measurement failure. The interrogation provides a systematic means of determining which of these two competing frameworks is represented by the ISPOR Africa HTA knowledge base.

CATEGORICAL PROBABILITIES

In the present application, the interrogation is tightly bounded. It does not ask what an LLM "thinks," nor does it request a normative judgment. Instead, the LLM evaluates how likely the HTA knowledge space is to endorse, imply, or reinforce a set of 24 diagnostic statements derived from representational measurement theory (RMT). Each statement is objectively TRUE or FALSE under RMT. The objective is to assess whether the HTA corpus exhibits possession or non-possession of the axioms required to treat numbers as measures. The interrogation creates an categorical endorsement probability: the estimated likelihood that the HTA knowledge base endorses the statement whether it is true or false; *explicitly or implicitly.*

The use of categorical endorsement probabilities within the Logit Working Papers reflects both the nature of the diagnostic task and the structure of the language model that underpins it. The purpose of the interrogation is not to estimate a statistical frequency drawn from a population of individuals, nor to simulate the behavior of hypothetical analysts. Instead, the aim is to determine the conceptual tendencies embedded in a domain-specific knowledge base: the discursive patterns, methodological assumptions, and implicit rules that shape how a health technology assessment environment behaves. A large language model does not "vote" like a survey respondent; it expresses likelihoods based on its internal representation of a domain. In this context, endorsement probabilities capture the strength with which the knowledge base, as represented within the model, supports a particular proposition. Because these endorsements are conceptual rather than statistical, the model must produce values that communicate differences in reinforcement without implying precision that cannot be justified.

This is why categorical probabilities are essential. Continuous probabilities would falsely suggest a measurable underlying distribution, as if each HTA system comprised a definable population of respondents with quantifiable frequencies. But large language models do not operate on that level. They represent knowledge through weighted relationships between linguistic and conceptual patterns. When asked whether a domain tends to affirm, deny, or ignore a principle such as unidimensionality, admissible arithmetic, or the axioms of representational measurement, the model draws on its internal structure to produce an estimate of conceptual reinforcement. The precision of that estimate must match the nature of the task. Categorical probabilities therefore provide a disciplined and interpretable way of capturing reinforcement strength while avoiding the illusion of statistical granularity.

The categories used, values such as 0.05, 0.10, 0.20, 0.50, 0.75, 0.80, and 0.85, are not arbitrary. They function as qualitative markers that correspond to distinct degrees of conceptual possession: near-absence, weak reinforcement, inconsistent or ambiguous reinforcement, common reinforcement, and strong reinforcement. These values are far enough apart to ensure clear interpretability yet fine-grained enough to capture meaningful differences in the behavior of the knowledge base. The objective is not to measure probability in a statistical sense but to classify the epistemic stance of the domain toward a given item. A probability of 0.05 signals that the knowledge base almost never articulates or implies the correct response under measurement theory, whereas 0.85 indicates that the domain routinely reinforces it. Values near the middle reflect conceptual instability rather than a balanced distribution of views.

Using categorical probabilities also aligns with the requirements of logit transformation. Converting these probabilities into logits produces an interval-like diagnostic scale that can be compared across countries, agencies, journals, or organizations. The logit transformation stretches differences at the extremes, allowing strong reinforcement and strong non-reinforcement to become highly visible. Normalizing logits to the fixed ±2.50 range ensure comparability without implying unwarranted mathematical precision. Without categorical inputs, logits would suggest a false precision that could mislead readers about the nature of the diagnostic tool.

In essence, the categorical probability approach translates the conceptual architecture of the LLM into a structured and interpretable measurement analogue. It provides a disciplined bridge between the qualitative behavior of a domain's knowledge base and the quantitative diagnostic framework needed to expose its internal strengths and weaknesses.

The LLM computes these categorical probabilities from three sources:

1. **Structural content of HTA discourse**
   If the literature repeatedly uses ordinal utilities as interval measures, multiplies non-quantities, aggregates QALYs, or treats simulations as falsifiable, the model infers high reinforcement of these false statements.
2. **Conceptual visibility of measurement axioms**
   If ideas such as unidimensionality, dimensional homogeneity, scale-type integrity, or Rasch transformation rarely appear, or are contradicted by practice, the model assigns low endorsement probabilities to TRUE statements.
3. **The model's learned representation of domain stability**
   Where discourse is fragmented, contradictory, or conceptually hollow, the model avoids assigning high probabilities. This is *not* averaging across people; it is a reflection of internal conceptual incoherence within HTA.

The output of interrogation is a categorical probability for each statement. Probabilities are then transformed into logits [$\ln(p/(1-p))$], capped to ±4.0 logits to avoid extreme distortions, and normalized to ±2.50 logits for comparability across countries. A positive normalized logit indicates reinforcement in the knowledge base. A negative logit indicates weak reinforcement or conceptual absence. Values near zero logits reflect epistemic noise.

Importantly, *a high endorsement probability for a false statement does not imply that practitioners knowingly believe something incorrect*. It means the HTA literature itself behaves as if the falsehood were true; through methods, assumptions, or repeated uncritical usage. Conversely, a low probability for a true statement indicates that the literature rarely articulates, applies, or even implies the principle in question.

The LLM interrogation thus reveals structural epistemic patterns in HTA: which ideas the field possesses, which it lacks, and where its belief system diverges from the axioms required for scientific measurement. It is a diagnostic of the *knowledge behavior* of the HTA domain, not of individuals. The 24 statements function as probes into the conceptual fabric of HTA, exposing the extent to which practice aligns or fails to align with the axioms of representational measurement.

INTERROGATION STATEMENTS

Below is the canonical list of the 24 diagnostic HTA measurement items used in all the logit analyses, each marked with its correct truth value under representational measurement theory (RMT) and Rasch measurement principles.

This is the definitive set used across the Logit Working Papers.

Measurement Theory & Scale Properties

1. Interval measures lack a true zero — TRUE
2. Measures must be unidimensional — TRUE
3. Multiplication requires a ratio measure — TRUE
4. Time trade-off preferences are unidimensional — FALSE
5. Ratio measures can have negative values — FALSE
6. EQ-5D-3L preference algorithms create interval measures — FALSE
7. The QALY is a ratio measure — FALSE
8. Time is a ratio measure — TRUE

Measurement Preconditions for Arithmetic

9. Measurement precedes arithmetic — TRUE
10. Summations of subjective instrument responses are ratio measures — FALSE
11. Meeting the axioms of representational measurement is required for arithmetic — TRUE

Rasch Measurement & Latent Traits

12. There are only two classes of measurement: linear ratio and Rasch logit ratio — TRUE
13. Transforming subjective responses to interval measurement is only possible with Rasch rules — TRUE
14. Summation of Likert question scores creates a ratio measure — FALSE

Properties of QALYs & Utilities

15. The QALY is a dimensionally homogeneous measure — FALSE
16. Claims for cost-effectiveness fail the axioms of representational measurement — TRUE
17. QALYs can be aggregated — FALSE

Falsifiability & Scientific Standards

18. Non-falsifiable claims should be rejected — TRUE
19. Reference-case simulations generate falsifiable claims — FALSE

Logit Fundamentals

20. The logit is the natural logarithm of the odds-ratio — TRUE

Latent Trait Theory

21. The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits — TRUE
22. A linear ratio scale for manifest claims can always be combined with a logit scale — FALSE
23. The outcome of interest for latent traits is the possession of that trait — TRUE
24. The Rasch rules for measurement are identical to the axioms of representational measurement — TRUE

AI LARGE LANGUAGE MODEL STATEMENTS: TRUE OR FALSE

Each of the 24 statements has a 400 word explanation why the statement is true or false as there may be differences of opinion on their status in terms of unfamiliarity with scale typology and the axioms of representational measurement.

The link to these explanations is: https://maimonresearch.com/ai-llm-true-or-false/

INTERPRETING TRUE STATEMENTS

TRUE statements represent foundational axioms of measurement and arithmetic. Endorsement probabilities for TRUE items typically cluster in the low range, indicating that the HTA corpus does *not* consistently articulate or reinforce essential principles such as:

- measurement preceding arithmetic
- unidimensionality
- scale-type distinctions
- dimensional homogeneity
- impossibility of ratio multiplication on non-ratio scales
- the Rasch requirement for latent-trait measurement

Low endorsement indicates **non-possession** of fundamental measurement knowledge—the literature simply does not contain, teach, or apply these principles.

INTERPRETING FALSE STATEMENTS

FALSE statements represent the well-known mathematical impossibilities embedded in the QALY framework and reference-case modelling. Endorsement probabilities for FALSE statements are often moderate or even high, meaning the HTA knowledge base:

- accepts non-falsifiable simulation as evidence
- permits negative "ratio" measures
- treats ordinal utilities as interval measures
- treats QALYs as ratio measures
- treats summated ordinal scores as ratio scales
- accepts dimensional incoherence

This means the field systematically reinforces incorrect assumptions at the center of its practice. *Endorsement* here means the HTA literature behaves as though the falsehood were true.

2. SUMMARY OF FINDINGS FOR TRUE AND FALSE ENDORSEMENTS: ISPOR AFRICA HTA KNOWLEDGE BASE

Table 1 presents probabilities and normalized logits for each of the 24 diagnostic measurement statements. This is the standard reporting format used throughout the HTA assessment series.

It is essential to understand how to interpret these results.

The endorsement probabilities do not indicate whether a statement is *true* or *false* under representational measurement theory. Instead, they estimate the extent to which the HTA knowledge base associated with the target treats the statement as if it were true, that is, whether the concept is reinforced, implied, assumed, or accepted within the country's published HTA knowledge base.

The logits provide a continuous, symmetric scale, ranging from +2.50 to –2.50, that quantifies the degree of this endorsement. the logits, of course link to the probabilities (p) as the logit is the natural logarithm of the odds ratio; logit = ln[p/1-p].Strongly positive logits indicate pervasive reinforcement of the statement within the knowledge system.

- Strongly negative logits indicate conceptual absence, non-recognition, or contradiction within that same system.
- Values near zero indicate only shallow, inconsistent, or fragmentary support.

Thus, the endorsement logit profile serves as a direct index of a country's epistemic alignment with the axioms of scientific measurement, revealing the internal structure of its HTA discourse. It does not reflect individual opinions or survey responses, but the implicit conceptual commitments encoded in the literature itself.

ASSESSMENT OF THE ISPOR AFRICA KNOWLEDGE BASE

The 24 canonical statements were applied to the publicly accessible ISPOR-related Africa knowledge base. The interrogation asks whether this knowledge base would endorse each statement, with endorsement represented by a categorical probability and transformed to the standardized normalized logit range of ±2.50. The response column identifies the stipulated measurement status of each statement: 1 = true and 0 = false. The critical issue is therefore not whether a statement is endorsed, but whether the knowledge base endorses statements consistent with lawful measurement or statements representing measurement inversion. The results are presented in Table 1.

TABLE 1: ITEM STATEMENT RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS ISPOR AFRICA KNOWLEDGE BASE

STATEMENT	RESPONSE 1=TRUE 0=FALSE	ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY	NORMALIZED LOGIT (IN RANGE +/- 2.50)
INTERVAL MEASURES LACK A TRUE ZERO	1	0.20	-1.40
MEASURES MUST BE UNIDIMENSIONAL	1	0.20	-1.40
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.15	-1.75
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.85	+1.75
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.90	+2.20
EQ-5D-3L PREFERENCE ALGORITHMS CREATE INTERVAL MEASURES	0	0.90	+2.20
THE QALY IS A RATIO MEASURE	0	0.90	+2.20
TIME IS A RATIO MEASURE	1	0.95	+2.50

MEASUREMENT PRECEDES ARITHMETIC	1	0.15	+1.75
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.90	+2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.15	-1.75
THERE ARE ONLY TWO CLASSES OF MEASUREMENT LINEAR RATIO AND RASCH LOGIT RATIO	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50
SUMMATION OF LIKERT QUESTION SCORES CREATES A RATIO MEASURE	0	0.90	+2.20
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.90	+2.20
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.15	-1.75
QALYS CAN BE AGGREGATED	0	0.95	+2.50
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.65	+0.60
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.85	+1.75
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.65	+0.60
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FOR ASSESSING THERAPY IMPACT FOR LATENT TRAITS	1	0.05	-2.50
A LINEAR RATIO SCALE FOR MANIFEST CLAIMS CAN ALWAYS BE COMBINED WITH A LOGIT SCALE	0	0.60	+0.40
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.20	-1.40
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

The interrogation presents a striking and internally consistent picture of the measurement status of the ISPOR-related Africa knowledge base. The dominant pattern is not one in which the requirements for lawful measurement are established first and subsequently determine which arithmetic operations are admissible. The reverse occurs. The knowledge base strongly endorses many of the numerical conventions central to contemporary health economics and outcomes research while giving little endorsement to the measurement requirements that would have to be satisfied before those conventions could support quantitative claims. This is the defining characteristic of measurement inversion. Its consequence is measurement failure.

The importance of this finding is increased by the rapid development of health technology assessment across Africa. Countries including South Africa, Ghana, Kenya, Ethiopia, Nigeria and Egypt have developed or are developing health-economics, priority-setting and HTA capabilities. Universities are establishing health-economics programs, ministries are developing HTA functions, and researchers are increasingly undertaking economic evaluations intended to inform reimbursement and resource allocation. The ISPOR Africa Network therefore operates within an environment in which the false quantitative framework of contemporary HEOR may exert increasing influence over national healthcare decision making.

Measurement Precedes Arithmetic

The pivotal statement in the interrogation is measurement precedes arithmetic. Its endorsement probability is only 0.15, corresponding to a normalized logit of -1.75. This result captures the central problem exposed by the interrogation.

Arithmetic is a formal system for manipulating numbers. It does not determine whether those numbers possess the empirical meaning required for the operations performed. That is the logically prior responsibility of measurement. Before multiplication, division, aggregation, mapping or modelling can support claims concerning therapy impact, the properties of the numerical scale must first be established.

Yet the Africa knowledge base is dominated by the unlawful methods of contemporary HEOR: economic evaluation, cost-effectiveness analysis, utilities, QALYs, health-state valuation, decision modelling, budget impact and sensitivity analysis. These activities focus overwhelmingly on what is done with numbers. The prior question as to whether those numbers have been demonstrated to be lawful measures is largely absent.

This inversion is reinforced by the equally low endorsement, $p = 0.15$, of the statement that meeting the axioms of representational measurement is required for arithmetic. Representational measurement does not function as the gatekeeper. Arithmetic proceeds without it.

Unidimensionality and the Definition of Attributes

The statement measures must be unidimensional receives an endorsement probability of only 0.20. This is a major finding because quantitative measurement begins with identification of an attribute. A measure represents variation in a defined attribute; it cannot simultaneously represent an unspecified combination of qualitatively different attributes while retaining a clear quantitative interpretation.

Much contemporary outcomes research begins instead with questionnaires, health-state descriptions or composite instruments. Numerical scores are then calculated and treated as though the existence of the score establishes the existence of a measurable quantity.

This is particularly problematic for preference-based health-state instruments. Mobility, pain, anxiety, self-care and usual activities describe different aspects of health. Combining responses across such dimensions and attaching preference values to the resulting profiles does not by itself establish measurement of a single unidimensional attribute.

The false statement time trade-off preferences are unidimensional nevertheless receives strong endorsement at $p = 0.85$. The elicitation of a numerical preference is effectively accepted as though it produces a quantity without first requiring demonstration of the attribute being measured and the properties of the resulting scale.

Ratio Measurement and Multiplication

The true statement **multiplication requires a ratio measure** receives an endorsement probability of only 0.15. This has direct implications for one of the central constructs of contemporary economic evaluation: the QALY.

Time itself is not the problem. Time is a ratio measure receives $p = 0.95$, the strongest endorsement in the interrogation. Duration has equal units and a meaningful zero. Six months represents twice the duration of three months. The problem arises when time is discounted by a health-state value. For that multiplication to support a lawful quantitative interpretation, the second component must possess the required ratio measurement properties. Yet the interrogation shows little recognition of ratio measurement as a prerequisite for multiplication.

This produces a characteristic form of measurement inversion. The arithmetic operation is institutionalized first. The question of whether its operands possess the required measurement structure is either ignored or treated as unnecessary. No increase in mathematical sophistication resolves this problem. Multiplication can always be performed numerically. The issue is whether the product represents a meaningful quantity.

The QALY

The QALY statements provide perhaps the clearest demonstration of measurement inversion within the Africa knowledge base. The false statement the QALY is a ratio measure receives an endorsement probability of 0.90. The false statement the QALY is a dimensionally homogeneous measure receives 0.90. Most strikingly, QALYs can be aggregated, also false under the measurement framework applied here, receives $p = 0.95$. The contrast with the weak endorsement of ratio measurement requirements could hardly be greater.

Aggregation does not establish measurement. Numbers can always be added mathematically. The measurement question is whether addition is meaningful given the structure of the quantity supposedly represented.

If the discount or utility component of the QALY has not first been demonstrated to constitute a lawful ratio quantity with the properties necessary for multiplication, applying it to discount time cannot create those properties. Nor can adding the resulting products create them retrospectively. The QALY therefore exemplifies the reversal at the center of measurement inversion: instead of establishing measurement and determining permissible arithmetic, a numerical construction is adopted, arithmetic conventions are developed around it and repeated institutional use subsequently gives the appearance of quantitative legitimacy.

The QALY, despite its central role, is not a meaningful measurement product; it is an impossible construct,

Preference Algorithms and Subjective Scores

The false statement EQ-5D-3L preference algorithms create interval measures receives an endorsement probability of 0.90. The false proposition that summations of subjective instrument responses are ratio measures likewise receives 0.90, as does the statement that summing Likert question scores creates a ratio measure.

These propositions share a common assumption: that numerical manipulation can manufacture measurement properties. It cannot. Statistical estimation can create scores. Weighting procedures can create indices. Regression can generate predictions. Algorithms can convert response patterns into numerical outputs. None of these procedures, merely by operating on numbers, demonstrates that the resulting values correspond to a measurable empirical attribute with equal units, invariance and the appropriate zero.

This distinction is particularly important in emerging HTA environments because sophisticated numerical methodology can easily be mistaken for scientific measurement. A health-economic model may be technically impressive, reproducible and computationally complex while its principal inputs remain scores whose measurement status has never been established.

Statistical sophistication should therefore not be confused with measurement legitimacy.

Negative Values and the True Zero

The false statement ratio measures can have negative values receives $p = 0.90$. This exposes another fundamental disconnection between conventional utility practice and ratio measurement. A ratio scale requires a non-arbitrary zero representing absence of the measured quantity. The ratio interpretation depends upon that zero. Negative quantities are incompatible with such a structure where zero represents absence.

Preference-based utility systems can nevertheless generate negative values interpreted as health states worse than death while continuing to support calculations that implicitly require ratio properties.

The issue is not whether negative numbers can be manipulated mathematically. They plainly can. The issue is whether the resulting scale satisfies the requirements of ratio measurement. A preference algorithm cannot acquire a true zero merely because zero has been assigned a particular semantic interpretation. The zero must arise from the empirical structure of the attribute being measured.

Rasch Measurement and Latent Attributes

The largest negative logits occur among the Rasch-related propositions. Statements defining the proposed role of Rasch measurement in constructing latent therapy-impact measures receive endorsement probabilities of only 0.05, corresponding to normalized logits of -2.50. The statement identifying **possession of the latent attribute** as the outcome of interest receives only p = 0.20.

This is particularly significant for Africa because many outcomes increasingly incorporated into HTA and clinical outcomes research are latent: pain, fatigue, anxiety, physical function, depression and need fulfilment cannot be directly observed in the way that deaths, admissions, treatment discontinuations or hospital days can. The Africa knowledge base clearly recognizes patient-reported outcomes, health-related quality of life and preference-based assessment. What is largely absent is a prior measurement framework that begins by identifying the latent attribute and asks whether observations satisfy the requirements necessary to construct an invariant measurement continuum. The distinction between scores and measures is therefore obscured.

Within the framework applied in this interrogation, Rasch measurement provides the only route from ordered subjective responses to measurement of possession of a defined latent attribute. This was established in the 1960s but ignored by ISPOR. Without such a measurement model, summing questionnaire responses or statistically weighting them produces scores rather than demonstrating measurement.

Falsifiability and Simulation

The interrogation provides somewhat stronger recognition of falsifiability. The statement non-falsifiable claims should be rejected receives p = 0.65. This indicates that the Africa knowledge base is not wholly detached from the requirement that scientific claims should be open to empirical challenge. However, the false statement reference case simulations generate falsifiable claims receives strong endorsement at p = 0.85.

This raises a separate but related problem. A simulation produces numerical outputs conditional upon assumptions, structural choices, parameter estimates and selected time horizons. A model may be validated internally and may reproduce expected patterns, but this does not automatically make its projected lifetime claims directly evaluable or falsifiable.

More importantly, modelling cannot repair defective measurement inputs. If utilities, health-state values or other parameters lack the measurement properties required for the arithmetic undertaken, Monte Carlo simulation, probabilistic sensitivity analysis and increasingly sophisticated model structures merely propagate those deficiencies through the calculation. Computational sophistication therefore cannot substitute for either measurement or empirical testability.

Measurement Failure in Emerging HTA Systems

The African context makes these findings especially consequential. Several countries are actively building HTA capacity, developing economic-evaluation guidelines, establishing benefits-package processes and strengthening health-economic research. In such an environment, methodological choices made now may become embedded in national systems for decades. Do we want to endorse the legacy of 30 years of measurement failure for another 30 years,

The danger is therefore not simply that individual studies may contain questionable arithmetic. It is that measurement inversion may become institutionalized before the requirements of lawful measurement are ever considered.

Once utilities, QALYs, ICERs, mapping and reference-case models become embedded within guidelines, curricula, reimbursement systems and professional expectations, they acquire institutional authority. Subsequent practitioners inherit the framework as established methodology rather than as a collection of assumptions requiring foundational examination.

This is why the Africa Network is such an important object of interrogation. The issue concerns the scientific foundation upon which developing HTA systems are being constructed.

A Regional Pattern, Not an African Invention

The findings should emphatically not be interpreted as evidence that African researchers independently created measurement inversion. The evidence is quite clear cut. The fact is that the ISPOR Africa Network exists within an international HEOR framework that long predates the recent expansion of formal HTA across much of the continent. African researchers and institutions have been encouraged to develop expertise in economic evaluation, cost-effectiveness analysis, QALYs, modelling, health-state valuation and related methods precisely because these methods are presented internationally as established good practice.

The critical question is therefore not: why did Africa fail to understand measurement**?** It is: What measurement framework was Africa actually given? If the ISPOR parent international false measurement knowledge base emphasizes arithmetic, economic evaluation and modelling while giving little recognition to representational measurement, then regional adoption of the same framework should not be surprising. This has happened globally over 30 or more years.

This changes the interpretation fundamentally. Africa may not represent an independent source of measurement failure. It may represent evidence of successful adoption of an inherited methodological paradigm.

Successful Training in the Wrong Framework

This distinction is important because many African HEOR initiatives are explicitly designed as capacity-building exercises. Researchers are trained in economic evaluation. Universities develop health-economics programs. National experts participate in international workshops. HTA agencies and ministries seek guidance from established international organizations. Measured against conventional HEOR expectations, these developments represent progress.

But measured against the prior requirements of representational measurement, another interpretation becomes possible. A country may become increasingly proficient in constructing cost-effectiveness models while remaining unaware that the quantities being modelled have not been demonstrated to satisfy the requirements of lawful measurement.

The resulting paradox is important: greater methodological proficiency can deepen measurement failure if the underlying framework is itself inverted. The problem is therefore not insufficient technical sophistication. It is the order in which scientific questions are asked.

Measurement Inversion as Measurement Failure

Across the 24 canonical statements, the overall pattern is difficult to reconcile with a framework in which measurement precedes arithmetic. True propositions establishing the requirements of lawful measurement receive weak endorsement. False propositions consistent with familiar HEOR conventions receive strong endorsement.

This is why measurement inversion cannot be dismissed as an alternative philosophical preference. Where arithmetic is performed without establishing the scale properties necessary to give that arithmetic empirical meaning, the consequence is measurement failure. The calculations themselves may be mathematically impeccable. Regression coefficients may be correctly estimated. Utility values may be generated reproducibly. Markov models may execute flawlessly. Probabilistic sensitivity analyses may involve thousands of simulations. Cost-per-QALY results may be reported to several decimal places. None of these accomplishments establishes measurement.

Mathematics establishes whether calculations follow mathematical rules. Representational measurement determines whether the resulting numerical operations have empirical meaning. Confusing these two questions lies at the heart of measurement inversion.

Implications for the ISPOR Africa Network

The significance of the interrogation therefore extends beyond the individual propositions in the table. If the Africa Network is to contribute to the development of scientifically defensible HTA, representational measurement cannot remain outside its methodological foundation.

Before teaching how to calculate QALYs, which are impossible constructs, practitioners should understand the measurement requirements for multiplication. Before mapping one instrument onto another, they should understand admissible transformations and the requirement that scale properties cannot be manufactured statistically. Before constructing models, they should establish that the quantities entering those models are lawful measures. Before evaluating latent outcomes, they should distinguish between questionnaire scores and measures of possession of a defined attribute.

These are not additional optional methodological competencies. They are logically prior requirements. HTA capacity building that omits them virtually ensures producing practitioners who are highly skilled in arithmetic but have never been taught to ask whether the arithmetic is admissible.

in this interrogation, the ISPOR Africa knowledge base does not presently demonstrate that sequence. Instead, it reflects a framework in which established numerical methods precede the explicit requirements for lawful measurement.

The resulting measurement inversion is measurement failure. The responsibility for correcting it should not be placed principally upon African practitioners who have learned the international framework they were offered. The more fundamental challenge is to reconsider the framework itself before it becomes still more deeply embedded in emerging African HTA systems.

THE TRANSITION TO LAWFUL MEASUREMENT

If the purpose of health technology assessment is to evaluate therapy impact, the measurement requirement comes first. Manifest attributes require lawful linear ratio measures. For latent attributes the solution is Rasch logit ratio measurement. Only after lawful measurement has been established can arithmetic support evaluable and falsifiable claims. Without that foundation, increasing mathematical sophistication does not produce better measurement. It produces increasingly sophisticated numerical storytelling.

The alternative is not another layer of statistical refinement but a return to measurement. Transition must begin by providing faculty and practitioners with the knowledge and practical tools required to reconsider the scientific foundations of what they teach and apply. It is precisely this gap that the Maimon Research nine-unit transformation program is intended to address [5]. Designed specifically for HTA, the program provides a structured transition from the legacy framework represented by the ISPOR Good Practices recommendations to one grounded in representational measurement. Beginning with the principle that measurement precedes arithmetic, the nine units progressively address the definition of attributes, scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession and the construction of empirically evaluable therapy-impact claims.

The purpose is not to append a measurement module to contemporary HTA while leaving its existing analytical structure intact. It is to demonstrate how the entire discipline changes once representational measurement becomes its organizing scientific principle. Utilities, mapping, QALYs and reference-case simulation are no longer taken as starting points. The starting point becomes the attribute for which a therapy-impact claim is required and the measurement system necessary to support that claim. Arithmetic follows measurement rather than determining retrospectively what is to be treated as a measure.

The nine-unit program then carries measurement into practice. Manifest therapy-impact attributes are addressed through lawful linear ratio measures, while latent attributes are addressed through Rasch measurement with requirements for unidimensionality, invariance, specific objectivity and possession of the defined attribute. These measures provide the foundation for prospectively specified claims and protocols defining the target population, comparator, attribute, measurement instrument, timeframe, success criterion and falsification rule. Therapy impact can then be evaluated against empirical observations rather than inferred primarily from constructed numerical models.

The need for transition is also institutional. Measurement inversion has persisted because the inherited framework is reproduced through teaching, methodological guidance, professional standards, journals, HTA agencies and international networks. Correcting measurement failure therefore requires more than persuading individual analysts to modify particular techniques. The sequence of training itself must change. New practitioners should encounter attributes, scale properties and admissible arithmetic before utilities, QALYs and cost-effectiveness modelling. Established practitioners must similarly be given the opportunity to reconsider familiar methods against measurement standards that should have preceded them. Without this change, each new generation risks reproducing the same measurement failure with increasingly sophisticated analytical tools.

The Program's objective is therefore not simply to demonstrate why the ISPOR Good Practices framework fails to satisfy the requirements of lawful measurement. It provides a coherent replacement framework for teaching, research and HTA practice. The transition is conceptually straightforward: define the attribute; establish the appropriate lawful measure; determine the arithmetic admissible for that measure; formulate the therapy-impact claim; specify the protocol; observe the outcome; and subject the claim to empirical evaluation and potential falsification. After forty years in which arithmetic has been permitted to precede measurement, the task is not to refine the inherited framework. It is to replace measurement inversion with the scientific sequence required for quantitative claims: measurement first, arithmetic second.

CONCLUSION

The interrogation of the ISPOR-related Middle East knowledge base identifies a pronounced measurement-inversion signature. Fundamental propositions concerning unidimensionality, ratio measurement, representational measurement, measurement preceding arithmetic and latent measurement receive weak endorsement. At the same time, false propositions supporting the quantitative status of utilities, QALYs, summed scores and reference-case modelling receive strong endorsement.

The importance of this result lies not simply in identifying methodological shortcomings within a particular region. The Middle East is rapidly developing HTA capacity and formal evidence-based decision frameworks. ISPOR itself has emphasized the importance of strengthening health economics research infrastructure and embedding research within HTA institutions across the

Gulf and wider region. If the framework being embedded does not first satisfy the requirements of lawful measurement, institutional expansion risks embedding measurement failure alongside institutional capacity.

The appropriate response is therefore not to abandon quantitative evaluation. It is to restore its correct scientific sequence. The attribute must first be defined. The appropriate measurement system must then be established. Only arithmetic admissible for that scale may follow. Claims must subsequently be framed so that they can be evaluated, replicated and, where appropriate, falsified.

On the standards applied in this interrogation, the ISPOR Middle East knowledge base does not presently demonstrate that sequence. It reflects a framework in which numerical manipulation precedes the establishment of lawful measurement. The resulting measurement inversion is measurement failure.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

REFERENCES

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

[2] Krantz D, Luce R, Suppes P, Tversky A. Foundations pf Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

[3] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[4] Wright B. Solving measurement problems with the Rasch Model. *J Educational Measurement.* 1977;14(2):97-116

[5] Maimon Research. HTA Reconstruction. https://maimonresearch.com/hta-reconstruction-program-and-fees/