MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

UNITED STATES: THE NATIONAL ENDORSEMENT OF FALSE MEASUREMENT CLAIMS AND MEASUREMENT INVERSION IN HEALTH TECHNOLOGY ASSESSMENT

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

LOGIT WORKING PAPER No 1888 AUGUST 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF MEASUREMENT INVERSION

ABSTRACT

Health technology assessment (HTA) and health economics and outcomes research (HEOR) occupy a central role in United States healthcare decision-making, supporting formulary selection, reimbursement, pricing and resource-allocation decisions through economic evaluation, cost-effectiveness analysis, patient-reported outcomes and decision modelling. The scientific legitimacy of these activities depends upon a prior requirement that is independent of economics or statistical sophistication: the quantities entering quantitative analyses must satisfy the requirements of lawful measurement. Representational measurement establishes these requirements through the principles that measurement must precede arithmetic, that arithmetic is constrained by scale properties, and that claims of therapy impact require either lawful linear ratio measurement for manifest attributes or Rasch logit ratio measurement for latent attributes.

This paper interrogates the United States HTA/HEOR knowledge base as a national scientific enterprise using twenty-four canonical statements derived from representational measurement. Rather than evaluating individual institutions, the interrogation examines the collective methodological framework represented across schools of pharmacy, schools of public health, medical schools, university research centers, professional organizations, journals and associated educational and research activities. Categorical endorsement probabilities and normalized logits are used to estimate the extent to which the national knowledge base reinforces either the principles of lawful measurement or the propositions that sustain measurement inversion. A companion program-content assessment evaluates whether representational measurement is explicitly represented within publicly available graduate teaching and research programs.

The interrogation demonstrates a consistent national pattern of measurement inversion. Foundational propositions concerning measurement preceding arithmetic, ratio measurement, unidimensionality, admissible arithmetic, dimensional homogeneity and Rasch measurement receive uniformly weak endorsement, while false propositions supporting utilities, QALYs, preference algorithms, summated ordinal scores, cost-effectiveness analysis and reference-case simulation receive strong endorsement. The companion program-content assessment reaches the same conclusion. Representational measurement is effectively absent from the publicly represented United States HTA/HEOR educational framework, whereas the conventional analytical methods of contemporary HTA are comprehensively represented. Together, these findings indicate that arithmetic and modelling are systematically taught and applied before the scientific conditions governing their admissibility have been established.

The implications extend beyond methodology to the scientific status of contemporary HTA itself. Once representational measurement is accepted as the governing framework for quantitative claims, utilities, QALYs, incremental cost-effectiveness ratios and reference-case simulation must first demonstrate conformity with the requirements of lawful measurement before they can support quantitative inference. The paper concludes that the United States HTA/HEOR knowledge base reproduces a nationally institutionalized framework of measurement inversion and argues that reconstruction around representational measurement, lawful ratio measurement and Rasch logit ratio measurement provides the only scientifically coherent foundation for future HTA research, education and healthcare decision-making.

INTRODUCTION

Health technology assessment (HTA) presents itself as a quantitative discipline. It seeks to generate scientifically credible claims concerning the comparative value of pharmaceutical products, medical devices and healthcare interventions through economic evaluation, cost-effectiveness analysis, cost-utility analysis, patient-reported outcomes and increasingly sophisticated decision models. These claims influence formulary decisions, reimbursement, pricing, clinical practice and the allocation of healthcare resources throughout the United States. Their scientific legitimacy, however, depends upon a prior condition that is independent of economics, statistics or modelling. The quantities entering these analyses must satisfy the requirements of lawful measurement. Measurement must precede arithmetic. Unless the quantities being manipulated possess the measurement properties required for multiplication, division and ratio formation, the resulting claims cannot be regarded as scientifically meaningful.

Representational measurement establishes these requirements. It defines the axioms governing quantitative measurement, the admissible arithmetic associated with different scales of measurement, the requirement for dimensional homogeneity and the distinction between manifest and latent attributes. For claims of therapy impact, representational measurement recognizes only two scientifically admissible quantitative measures: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. These are not competing methodological preferences. They are the scientific conditions that must be satisfied before arithmetic can support credible, empirically evaluable, independently replicable and potentially falsifiable claims.

Previous investigations have demonstrated that contemporary HTA has evolved along a markedly different trajectory. National interrogations of HTA agencies, professional organizations, university curricula, research centers and program-content knowledge bases in Australia, Canada, New Zealand and the United States, together with regional knowledge bases associated with ISPOR in Asia, Europe, Latin America, the Middle East and Africa, have consistently identified the same pattern. The principles of representational measurement receive weak endorsement, while propositions incompatible with lawful measurement receive strong endorsement. Complementary evaluations of curricula and publicly available program content reach the same conclusion. Representational measurement is largely absent from graduate education and research, while utilities, QALYs, preference elicitation, cost-effectiveness analysis, decision modelling and reference-case simulation dominate the analytical framework. The result is measurement inversion: arithmetic is accepted and applied before the scientific conditions governing its admissibility have been established.

The present investigation extends these earlier studies by interrogating the United States HTA/HEOR knowledge base as a national scientific entity. Rather than examining individual universities, research centers or states in isolation, the unit of analysis is the collective knowledge base through which HTA and health economics are taught, researched and applied across the United States. This national knowledge base comprises the publicly represented methodological framework embodied in schools of pharmacy, schools of public health, medical schools, university research centers, health economics and outcomes-research programs, professional organizations, journals, conferences and associated educational and research activities. It represents the cumulative intellectual environment that shapes successive generations of HTA practitioners.

The national perspective is important because it addresses a question that individual institutional or state interrogations cannot answer. If measurement inversion is observed repeatedly across California, Texas, Florida, Colorado, Utah, Oregon, Iowa and other states, does this reflect isolated institutional variation, or does it characterize the United States HTA/HEOR knowledge base itself? The national interrogation tests this proposition directly. Applying the same twenty-four canonical statements derived from representational measurement, it estimates categorical probabilities and normalized logits describing the extent to which the national knowledge base endorses either the scientific principles of lawful measurement or the propositions that sustain measurement inversion.

The resulting interrogation demonstrates a remarkably coherent national pattern. Across the United States HTA/HEOR knowledge base, the scientific principles governing quantitative measurement receive consistently weak endorsement, while propositions supporting utilities, QALYs, preference algorithms, cost-effectiveness analysis and reference-case modelling receive consistently strong endorsement. The defining role of ratio measurement in governing arithmetic has largely disappeared from the conceptual framework of contemporary HTA. In its place stands an analytical system whose quantitative claims depend upon treating numerical constructions as though they possessed measurement properties that have never been established. The consequence is not simply methodological weakness but the national endorsement of false measurement claims and the institutional reproduction of measurement inversion.

The objective of this paper is therefore broader than evaluating the methodological orientation of individual institutions or jurisdictions. It asks whether the United States HTA/HEOR knowledge base, viewed as a single national scientific enterprise, is organized around the accepted principles of representational measurement or whether it reproduces an inherited analytical framework that cannot satisfy the standards of normal quantitative science. The findings also identify a constructive path forward. Once representational measurement is restored as the organizing scientific principle, HTA acquires a coherent and scientifically defensible foundation based upon linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes. The national interrogation therefore provides both a diagnosis of the current scientific status of United States HTA and the framework for its reconstruction around the principles of lawful quantitative measurement.

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.

UNITED STATES: HEALTH TECHNOLOGY ASSESSMENT (HTA) AND HEALTH ECONOMICS AND OUTCOMES RESEARCH (HEOR) KNOWLEDGE BASE

The United States possesses the largest and most influential health technology assessment (HTA), health economics and outcomes research (HEOR) knowledge base in the world. Distributed across schools of pharmacy, schools of public health, medical schools, colleges of medicine, university research centers, health systems and affiliated research institutes, it encompasses an extensive network of teaching, research and methodological development devoted to comparative effectiveness research, economic evaluation, health services research, patient-reported outcomes, decision science, implementation science, real-world evidence, value assessment and healthcare policy. Collectively these activities shape graduate education, methodological innovation, reimbursement policy and healthcare decision-making throughout the United States while exerting substantial influence on HTA practice internationally.

The national knowledge base extends well beyond dedicated HTA research centers. It comprises university programs in health economics, pharmacoeconomics, comparative effectiveness research, outcomes research, value assessment, health policy, population health, epidemiology and quantitative evaluation. Major academic health systems, including leading public and private universities, contribute through graduate teaching, multidisciplinary research programs, methodological publications, seminars, workshops, professional education and collaborations with federal and state agencies, healthcare systems, pharmaceutical manufacturers, biotechnology companies and medical device organizations. Together these institutions form an integrated national knowledge base that defines the analytical framework through which successive generations of HTA and HEOR practitioners are educated and trained.

The methodological orientation of this national knowledge base is equally apparent. Publicly available descriptions of research priorities, graduate curricula, faculty expertise and research outputs demonstrate comprehensive engagement with the contemporary analytical framework of HTA and HEOR. Economic evaluation, cost-effectiveness analysis, cost-utility analysis, patient-reported outcomes, health utility measurement, preference elicitation, decision analysis, decision trees, Markov modelling, microsimulation, probabilistic sensitivity analysis, real-world evidence, comparative effectiveness research, value assessment and healthcare resource allocation are extensively represented. Advanced statistical methods, predictive analytics, artificial intelligence and increasingly sophisticated decision-analytic models are presented as central components of both graduate education and research. The United States therefore represents an exceptionally advanced and technically sophisticated HTA and HEOR environment whose methodological influence extends well beyond its national boundaries.

The present review, however, does not evaluate the technical sophistication, scientific productivity or policy influence of this research enterprise. Nor does it assess the competence of individual investigators or the quality of individual research contributions. The question addressed is considerably narrower but scientifically more fundamental. Does the United States HTA/HEOR knowledge base explicitly recognize representational measurement as the scientific foundation for quantitative claims before introducing the analytical methods of contemporary HTA? Specifically, does the national knowledge base establish the principles of measurement preceding arithmetic, scales of measurement, ratio measurement, admissible transformations, dimensional homogeneity, the distinction between manifest and latent attributes, linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes as the scientific framework governing subsequent quantitative analysis?

These questions are not matters of methodological preference. They determine whether the arithmetic operations performed throughout contemporary HTA satisfy the accepted requirements of lawful measurement. Utilities, QALYs, incremental cost-effectiveness ratios and reference-case simulation models all depend upon multiplication, division, aggregation and ratio formation. Unless the quantities entering these operations possess the measurement properties required by representational measurement, increasingly sophisticated statistical methods, machine learning algorithms and decision models cannot establish their scientific legitimacy. The critical issue is therefore not whether the United States teaches and applies contemporary HTA comprehensively—it clearly does—but whether the scientific conditions governing that quantitative framework are themselves explicitly represented throughout the national knowledge base.

The United States provides perhaps the most demanding test of this proposition. It contains the world's largest concentration of HTA, HEOR and health economics expertise, encompassing major academic programs and research centers across numerous states, together with extensive professional organizations, journals and educational networks. If representational measurement had become established anywhere as the organizing scientific framework for HTA, it would reasonably be expected to appear within this national knowledge base. The interrogation reported here therefore examines the United States HTA/HEOR knowledge base using the twenty-four canonical statements derived from representational measurement. The objective is to determine whether lawful measurement provides the scientific foundation for the United States HTA enterprise or whether, as previous state, national and international investigations suggest, the knowledge base continues to reproduce the four-decade legacy of measurement inversion that has come to characterize contemporary health technology assessment.

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
25.

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. ASSESSMENT OF THE EXTENT OF MEASUREMENT INVERSION

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.

TABLE 1: ITEM STATEMENT RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS MEASUREMENT INVERSION UNITED STATES

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	0.15	-1.75
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.00
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.90	+2.20
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	0.10	-2.00
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.90	+2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.10	-2.00
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	0.75	+1.10
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.85	+1.75
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.75	+1.10
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	0.15	-1.75
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF	1	0.05	-2.50

REPRESENTATIONAL MEASUREMENT			

ASSESSMENT

The interrogation of the United States HTA/HEOR knowledge base demonstrates a comprehensive and deeply institutionalized pattern of measurement inversion. The result is not a marginal imbalance between competing methodological traditions. It is a polarized epistemic structure in which the scientific propositions required to establish lawful quantitative measurement receive consistently weak endorsement, while propositions sustaining utilities, QALYs, cost-effectiveness analysis, preference algorithms and reference-case simulation receive consistently strong endorsement. The national knowledge base therefore privileges arithmetic, statistical manipulation and modelling before establishing that the quantities subjected to those operations possess the required measurement properties.

The contrast begins with the foundational propositions of representational measurement. Interval measures lack a true zero receives an endorsement probability of only 0.20, corresponding to a normalized logit of −1.40. Measures must be unidimensional receives 0.15 and −1.75. Multiplication requires a ratio measure receives only 0.10 and −2.00. Measurement precedes arithmetic is likewise endorsed at 0.10, with a normalized logit of −2.00, while meeting the axioms of representational measurement is required for arithmetic also receives 0.10 and −2.00. These are not peripheral propositions. Together they establish the scientific boundary between lawful measurement and the manipulation of numbers. Their weak endorsement indicates that the national HTA/HEOR knowledge base does not possess representational measurement as its organizing quantitative framework.

The marginally stronger endorsement of interval measures lacks a true zero at 0.20 and −1.40 does not alter this conclusion. Recognition of isolated features of scale theory is insufficient where the consequences for arithmetic are not carried through. A knowledge base may acknowledge that interval scales lack a true zero while still treating preference scores, utilities and health-state values as though they support multiplication, division and ratio formation. The decisive question is whether scale properties constrain arithmetic in practice. The endorsement profile demonstrates that they do not.

The national knowledge base gives exceptionally weak endorsement to the propositions governing latent-attribute measurement. There are only two classes of measurement, linear ratio and Rasch logit ratio, receives 0.05 and the minimum normalized logit of −2.50. Transforming subjective responses to interval measurement is only possible with Rasch rules also receives 0.05 and −2.50. The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits receives the same 0.05 and −2.50, as does the proposition that the Rasch rules for measurement are identical to the axioms of representational measurement. The outcome of interest for latent traits is the possession of that trait receives only 0.15 and −1.75.

Taken together, these results demonstrate near-total non-possession of the measurement framework required for latent attributes. Pain, fatigue, physical function, depression, anxiety, need fulfilment and other patient-reported constructs are central to contemporary outcomes research. Yet the national knowledge base gives almost no recognition to the requirement that such attributes must be defined unidimensionally, calibrated through person-item conjoint measurement, tested for invariance and interpreted in terms of possession of the latent attribute. In the absence of these requirements, summed scores, preference weights and algorithmically generated values are treated as though numerical assignment itself establishes measurement.

The weakness of the representational-measurement framework is mirrored by strong reinforcement of propositions incompatible with lawful measurement. Time trade-off preferences are unidimensional receives an endorsement probability of 0.90 and a normalized logit of +2.20. Ratio measures can have negative values receives 0.90 and +2.20. EQ-5D-3L preference algorithms create interval measures and the QALY is a ratio measure each receive 0.90 and +2.20. Summations of subjective instrument responses are ratio measures and summation of Likert question scores creates a ratio measure also receive 0.90 and +2.20.

This cluster identifies the central numerical assumptions sustaining contemporary HTA and HEOR. Time trade-off responses are treated as though they represent a single unidimensional attribute. Preference algorithms are treated as though they transform multidimensional ordinal classifications into interval or ratio quantities. Summed questionnaire responses are treated as though aggregation creates measurement. Negative values are accepted within quantities described or used as ratio measures. None of these assumptions satisfies the requirements of representational measurement, yet each is strongly embedded within the national knowledge base.

The endorsement of the QALY framework is especially decisive. The QALY is a ratio measure receives 0.90 and +2.20. The QALY is a dimensionally homogeneous measure receives the same 0.90 and +2.20. QALYs can be aggregated receives the highest endorsement category, 0.95 and +2.50. The national knowledge base therefore behaves as though the QALY possesses all the properties required for multiplication, aggregation and ratio formation, even though the propositions establishing those properties receive little or no endorsement.

This is measurement inversion in its clearest form. The knowledge base strongly reinforces the result of the arithmetic while weakly recognizing the scientific conditions required to justify the arithmetic. Utilities are multiplied by time without first demonstrating that utilities are ratio measures. QALYs are aggregated without establishing common units, dimensional homogeneity or invariance. Differences in cost are divided by differences in QALYs without demonstrating that the denominator is a lawful ratio quantity. Arithmetic is assumed to create quantitative meaning rather than being permitted only after measurement has been established.

The one strongly endorsed true proposition is time is a ratio measure, with a probability of 0.95 and a normalized logit of +2.50. This is unsurprising because time is a familiar manifest attribute with established units and a non-arbitrary zero for elapsed duration. Its strong endorsement, however, makes the failure of the QALY framework more rather than less apparent. The multiplication of time by another numerical value is lawful only if the second value also possesses the properties required for multiplication. The national knowledge base correctly recognizes time as a ratio measure but fails to impose the same standard on the utility component. The legitimacy of one factor cannot compensate for the measurement failure of the other.

Claims for cost-effectiveness fail the axioms of representational measurement receives only 0.15 and −1.75. This weak endorsement confirms that the consequences of the measurement failure are not recognized. If utilities are not lawful ratio measures, QALYs cannot be created through multiplication. If QALYs are not lawful ratio quantities, incremental cost-effectiveness ratios cannot represent meaningful quantitative ratios. The weak endorsement of this proposition indicates that the national knowledge base does not connect the measurement status of the underlying quantities to the legitimacy of cost-effectiveness claims.

The treatment of reference-case simulation reinforces the same pattern. Reference-case simulations generate falsifiable claims is a false statement under the framework applied here, yet it receives an endorsement probability of 0.85 and a normalized logit of +1.75. The knowledge base therefore strongly reinforces the assumption that model-generated outputs constitute empirically evaluable scientific claims. This confuses the internal operation of a model with the external testing of a proposition. A simulation may be internally consistent, probabilistically elaborate and supported by extensive sensitivity analysis, but those characteristics do not make its projected outputs falsifiable evidence.

Non-falsifiable claims should be rejected receives 0.75 and +1.10. This indicates moderate recognition of falsifiability as a general scientific standard. The contradiction is revealing. The national knowledge base recognizes, at least abstractly, that non-falsifiable claims should be rejected while simultaneously endorsing reference-case simulation as a source of falsifiable claims. The difficulty is therefore not a complete absence of the language of scientific evaluation. It is the failure to apply that language to the dominant methods of contemporary HTA.

The logit is the natural logarithm of the odds ratio also receives 0.75 and +1.10. This suggests that the mathematical definition of the logit is relatively familiar. Yet recognition of the formula does not imply possession of Rasch measurement. The national knowledge base may recognize logits statistically while failing to recognize their role in constructing invariant measures of latent attributes. The contrast between 0.75 and +1.10 for the mathematical definition of a logit and 0.05 and −2.50 for the Rasch logit ratio scale as the basis for latent therapy-impact claims demonstrates this separation clearly.

The proposition that a linear ratio scale for manifest claims can always be combined with a logit scale receives an endorsement probability of 0.60 and a normalized logit of +0.40. This relatively weak positive value indicates uncertainty or conceptual instability. The national knowledge base does not firmly recognize that manifest and latent attributes constitute distinct measurement systems whose outputs cannot automatically be combined merely because both are expressed numerically. Values near zero are particularly revealing because they indicate that the necessary distinction is neither clearly possessed nor clearly rejected.

The overall polarity of the results is unmistakable. Foundational true propositions concerning lawful measurement generally receive endorsement probabilities between 0.05 and 0.20, corresponding to normalized logits between −2.50 and −1.40. By contrast, false propositions supporting utilities, QALYs, summated scores, preference algorithms and reference-case modelling receive probabilities between 0.85 and 0.95, corresponding to normalized logits between +1.75 and +2.50. This is not an ambiguous or mixed profile. It is an inversion of the scientific order required for quantitative analysis.

The national significance of this finding is substantial. State-level interrogations can always be challenged as reflecting the peculiarities of particular universities, regional traditions or local institutional structures. The United States HTA/HEOR knowledge base removes that defense. It is constructed from the combined academic and research environments of multiple states, including major concentrations of expertise in California, Texas, Florida, Colorado, Utah, Oregon, Iowa and the other states to be incorporated. These environments differ greatly in size, institutional history, disciplinary composition and regional influence. Yet the same measurement profile is reproduced.

The national knowledge base therefore cannot be described as a collection of isolated methodological omissions. It represents a coherent epistemic system transmitted through schools of pharmacy, public health programs, medical schools, health-policy departments, health-services research centers, outcomes-research units, professional organizations, journals, conferences and industry training. Students and practitioners encounter the same analytical vocabulary: utilities, QALYs, cost-effectiveness thresholds, ICERs, decision trees, Markov models, microsimulation, probabilistic sensitivity analysis and value assessment. What they do not encounter first is the scientific framework required to determine whether the quantities entering those methods are lawful measures.

This is the institutionalization of measurement inversion at the national level. The United States HTA/HEOR knowledge base has developed extensive technical competence while remaining epistemically closed to the science of measurement. It has produced generations of analysts who can estimate utilities, calculate QALYs, construct models and report ICERs without first being required to establish unidimensionality, ratio properties, admissible transformations, dimensional homogeneity or invariance.

The finding should not be interpreted as an allegation against individual faculty members, programs or institutions. The knowledge base is the cumulative publicly represented framework of a discipline. Most practitioners have been educated within the same inherited methodology and reproduce the standards accepted by universities, journals, professional associations, healthcare organizations and industry. The pattern is institutional rather than personal. Its persistence reflects the absence of representational measurement from the educational sequence rather than deliberate rejection by individual investigators.

The implications, however, remain uncompromising. Contemporary HTA cannot claim quantitative legitimacy merely because its methods are widely taught, technically sophisticated or administratively useful. Scientific status depends upon lawful measurement. Once measurement precedes arithmetic, every construct must demonstrate the scale properties necessary for the operations performed. Utilities must demonstrate lawful ratio properties before multiplication. QALYs must demonstrate dimensional homogeneity and common units before aggregation. ICERs must contain lawful quantities before ratio formation. Modelled claims must be framed so that they can be subjected to empirical evaluation and potential falsification.

The United States HTA/HEOR knowledge base fails at this prior level. Its central quantitative constructs are accepted before their measurement legitimacy is established. Representational measurement is not simply missing from the knowledge base; its absence is the condition that allows the inherited framework to continue. Once the axioms of measurement are introduced, utilities, QALYs, ICERs and reference-case simulations cease to be unquestioned analytical standards and become objects of scientific evaluation.

The conclusion is therefore unavoidable. The United States HTA/HEOR knowledge base is organized around measurement inversion. It begins with arithmetic, modelling and value construction and assumes that measurement follows. Normal quantitative science requires the opposite sequence. The attribute must first be defined, the appropriate measurement system established, the permissible arithmetic determined and the resulting claim subjected to empirical evaluation, replication and falsification.

For therapy-impact claims, this reconstruction requires two measurement foundations. Manifest attributes require lawful linear ratio measurement. Latent attributes require Rasch logit ratio measurement, with unidimensionality, person-item conjoint measurement, invariance, specific objectivity and possession of the defined attribute. Only after those conditions have been satisfied can arithmetic operations support scientifically credible claims.

The national interrogation therefore identifies not a need for incremental methodological refinement but a need for reconstruction. The inherited HTA framework cannot be repaired by more elaborate modelling, additional sensitivity analysis, artificial intelligence or computational sophistication. Those developments occur after the measurement decision and cannot correct its failure. The required transformation is to restore measurement to its governing position before arithmetic.

Until that transformation occurs, the United States HTA/HEOR knowledge base should be recognized for what the endorsement profile demonstrates it to be: an extensive and technically accomplished system of numerical analysis built upon the institutionalization of measurement inversion rather than the accepted foundations of lawful quantitative science.

IRRECONCILABLE PROGRAM CONTENT

A companion assessment of the United States HTA/HEOR program-content knowledge base extends the present findings beyond measurement inversion to a direct examination of what is actually taught and represented throughout graduate education, university research centers and associated research and training programs across the United States. The results are unequivocal. The scientific foundations of representational measurement are effectively absent from the national program-content knowledge base. There is no explicit recognition of measurement preceding arithmetic, the axioms of representational measurement, scales of measurement as constraints on arithmetic, ratio measurement, admissible transformations, dimensional homogeneity, the distinction between manifest and latent attributes, linear ratio measurement or Rasch logit ratio measurement. In contrast, the contemporary HTA and HEOR framework is comprehensively represented through utilities, QALYs, cost-effectiveness analysis, preference elicitation, decision analysis, simulation modelling, reference-case methods and associated analytical techniques.

These two bodies of knowledge are not complementary; they are scientifically irreconcilable. Representational measurement establishes the criteria by which every quantitative construct must be judged. Once those criteria are introduced, the contemporary HTA framework cannot simply continue unchanged because its central constructs must first demonstrate that they satisfy the requirements for lawful measurement. The United States program-content review therefore demonstrates considerably more than an educational omission. It identifies a national educational and research framework structured around analytical methods whose continued acceptance depends upon the absence of representational measurement. The omission is not incidental; it is the condition that allows the contemporary HTA and HEOR framework to survive. Throughout the national knowledge base, arithmetic, modelling and numerical manipulation are taught before students and researchers are provided with the scientific principles required to determine whether the quantities entering those analyses possess the measurement properties necessary to support the operations performed. The result is the systematic reproduction of program-content inversion across the United States HTA and HEOR enterprise.

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. Latent attributes require Rasch logit ratio measurement. Only after lawful measurement has been established can arithmetic support credible, empirically evaluable and potentially 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. Across the United States, transition must begin by providing faculty, researchers and practitioners with the knowledge and practical tools required to reconsider the scientific foundations of what they teach, investigate 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 analytical framework represented by contemporary HTA and 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 discipline itself changes once representational measurement becomes its organizing scientific principle. Utilities, mapping, QALYs and reference-case simulation are no longer treated as the starting points for quantitative analysis. The starting point becomes the attribute for which a therapy-impact claim is required together with the lawful measurement system necessary to support that claim. Arithmetic follows measurement rather than retrospectively determining 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 measurement, while latent attributes are addressed through Rasch measurement with explicit requirements for unidimensionality, invariance, specific objectivity and possession of the defined attribute. These measures provide the foundation for prospectively specified claims and protocols identifying the target population, comparator, attribute, measurement instrument, observation period, 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 throughout the United States HTA and HEOR knowledge base through university curricula, graduate education, methodological guidance, professional organizations, journals, HTA agencies, research centers and national and international professional networks. Correcting measurement failure therefore requires considerably more than persuading individual investigators to modify particular analytical techniques. The sequence of education itself must change. Students should first encounter attributes, scales of measurement, admissible arithmetic and lawful ratio measurement before they are introduced to utilities, QALYs, cost-effectiveness analysis and decision modelling. Established practitioners must likewise be given the opportunity to reassess familiar methods against the measurement standards that should always have preceded them. Without this change, each successive generation will continue to reproduce the same measurement failure with increasingly sophisticated analytical tools.

The program's objective is therefore not simply to demonstrate why the legacy HTA 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 evaluation protocol; observe the outcome; and subject the claim to empirical evaluation, independent replication and potential falsification. After four decades 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.

The United States HTA and HEOR community now faces a professional responsibility that extends beyond methodological debate. The evidence documenting measurement inversion is no longer confined to isolated institutions or individual critiques. It has accumulated across national HTA agencies, professional organizations, university curricula, research centers, program-content evaluations and, now, the United States HTA/HEOR knowledge base itself. To continue teaching and applying the inherited framework without informing students, researchers and practitioners that its scientific foundations have been fundamentally challenged would be inconsistent with the responsibilities of a university and the norms of scientific inquiry. The appropriate response is neither to ignore nor suppress these findings, but to examine them openly and critically. Equally important, this is not a critique without a solution. A coherent alternative exists. Representational measurement provides the scientific foundation, and the Maimon Research nine-unit transition program provides a practical pathway by which universities, research centers and professional organizations can move from the legacy framework of measurement inversion to a discipline grounded in lawful measurement, empirically evaluable claims and the standards of normal quantitative science.

CONCLUSION

The interrogation of the United States HTA/HEOR knowledge base demonstrates the national endorsement of a four-decade legacy of false measurement claims and measurement inversion. The knowledge base gives consistently weak endorsement to the scientific principles of representational measurement while strongly endorsing propositions that sustain utilities, QALYs, cost-effectiveness analysis, preference algorithms and reference-case modelling. The central failure is the disappearance of ratio measurement as the governing principle of quantitative analysis. Once it is recognized that multiplication, division and ratio formation require lawful ratio measures, the analytical structure of contemporary HTA can no longer be sustained. Increasingly sophisticated modelling, artificial intelligence, machine learning and computational refinement cannot compensate for the absence of lawful measurement.

The companion program-content assessment reaches precisely the same conclusion from a different direction. Representational measurement, scales of measurement, admissible arithmetic, dimensional homogeneity, linear ratio measurement and Rasch measurement are absent from the publicly represented United States HTA and HEOR curriculum, while the methods of contemporary HTA are comprehensively represented. This is not simply an educational imbalance but an irreconcilable scientific contradiction. The contemporary HTA framework can survive only while representational measurement remains absent because, once introduced, its standards immediately expose the incompatibility of utilities, QALYs, cost-effectiveness ratios and reference-case modelling with lawful measurement. The nationwide convergence of the measurement inversion and program-content interrogations demonstrates that this is not an isolated institutional problem. It characterizes the United States HTA/HEOR knowledge base as a whole.

The implications are unavoidable. Contemporary HTA, as presently constituted, should no longer be taught as a graduate quantitative discipline without first confronting the scientific requirements governing quantitative measurement. Universities have a responsibility to ensure that quantitative subjects satisfy the accepted standards of representational measurement before teaching the arithmetic that depends upon those standards. That responsibility extends equally to schools of pharmacy, schools of public health, medical schools, health economics programs, research centers, professional organizations and the journals through which the discipline reproduces itself. Students and practitioners should not be introduced to utilities, QALYs, cost-effectiveness ratios and decision models until they first understand the measurement principles that determine whether those constructs are scientifically admissible.

The United States now has an opportunity to replace a forty-year legacy of measurement inversion with a scientifically coherent foundation for health technology assessment. Representational measurement resolves the demarcation problem by restoring measurement before arithmetic and limiting claims of therapy impact to the only two admissible measures: lawful linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. This is not another methodological option within contemporary HTA, nor is it an incremental refinement of the existing framework. It is the only analytical framework consistent with the accepted standards of normal quantitative science.

The evidence presented in this and the companion investigations also creates a broader professional obligation. The critique of contemporary HTA can no longer be dismissed as an isolated theoretical concern. Independent interrogations of national agencies, university curricula, research centers, program-content knowledge bases and now the United States HTA/HEOR knowledge base have consistently reached the same conclusion. Once such evidence exists, universities, educators, journal editors, professional organizations and HTA agencies have a duty of care to ensure that students, researchers and practitioners are informed that the scientific foundations of the inherited framework have been fundamentally challenged. Equally important, they should also be informed that a viable scientific alternative exists. Representational measurement and the associated transition framework provide a practical pathway for reconstructing HTA around lawful measurement, admissible arithmetic and empirically evaluable and falsifiable claims. The challenge facing the United States is therefore no longer whether measurement inversion exists. It is whether the national HTA and HEOR community is prepared to acknowledge the evidence and begin the scientific transition that normal quantitative science now requires.

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/