MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

CALIFORNIA: THE ENDORSEMENT OF MEASUREMENT INVERSION IN HEALTH TECHNOLOGY ASSESSMENT

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

LOGIT WORKING PAPER No 1854 AUGUST 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF MEASUREMENT INVERSION

ABSTRACT

Health technology assessment (HTA) is founded upon quantitative claims concerning the comparative value of pharmaceutical products, medical devices and healthcare interventions. The scientific legitimacy of these claims depends upon a prior requirement: measurement must precede arithmetic. Representational measurement theory establishes the conditions under which arithmetic is admissible, distinguishing between nominal, ordinal, interval and ratio scales, requiring dimensional homogeneity, and recognizing only two lawful approaches to therapy-impact measurement: linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes. This paper examines whether these scientific principles are explicitly recognized within the California HTA knowledge base.

California was selected because it contains one of the largest and most influential concentrations of HTA, health economics and outcomes research expertise in the United States. The assessment forms the first in a planned series of state-level interrogations designed to determine whether representational measurement or measurement inversion provides the conceptual foundation for graduate education, research and methodological development. Using twenty-four canonical statements derived from representational measurement theory, a large language model interrogation generated categorical endorsement probabilities and normalized logits describing the extent to which the California knowledge base reinforces true and false propositions concerning lawful measurement.

The results identify a coherent pattern of measurement inversion. Fundamental propositions concerning representational measurement, ratio measurement, measurement preceding arithmetic, admissible arithmetic, dimensional homogeneity and Rasch measurement receive consistently weak endorsement. In contrast, false propositions supporting utilities, QALYs, preference algorithms, summed ordinal scores, aggregation and reference-case modelling receive strong endorsement. The defining scientific role of ratio measurement in supporting multiplication, division and ratio formation is effectively absent from the knowledge base. A companion program-content review reaches the same conclusion, demonstrating that the principles of representational measurement are largely absent from graduate HTA curricula while the contemporary HTA framework is comprehensively represented.

The findings indicate that California has inherited and institutionalized a thirty-year international legacy of measurement inversion rather than establishing lawful measurement as the foundation *for quantitative claims. The consequences extend beyond methodological preference to the scientific status of the discipline itself. Once the standards of representational measurement are applied, the analytical framework of contemporary HTA cannot be sustained because its central constructs fail the measurement conditions required for their arithmetic. The paper concludes that California has an opportunity to abandon this legacy and establish a new scientific foundation for HTA based upon representational measurement, with therapy-impact claims restricted to linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes.*

INTRODUCTION

Health technology assessment (HTA) presents itself as a quantitative discipline. It supports 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 and the allocation of healthcare resources. The scientific legitimacy of these activities depends upon a single prior condition: that the quantities entering the analysis 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 identifies only two scientifically admissible measures: **linear ratio measures** for manifest attributes and **Rasch logit ratio measures** for latent attributes. These are not methodological preferences. They represent the scientific conditions that must be satisfied before arithmetic can support credible, empirically evaluable and potentially falsifiable claims.

A growing body of evidence suggests that contemporary HTA has developed along a different trajectory. Previous interrogations of national HTA agencies, professional organizations, university curricula, research centers and program-content knowledge bases in Australia, New Zealand, Canada 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 lawful measurement receive weak endorsement, while propositions incompatible with representational measurement receive strong endorsement. Reviews of publicly available curricula and program content have reached the same conclusion. Representational measurement is largely absent, while utilities, QALYs, preference elicitation, cost-effectiveness analysis and reference-case modelling dominate teaching and research. The result is measurement inversion: arithmetic is taught and applied before the scientific conditions governing its admissibility have been established.

The present paper begins a new series of state-level interrogations of the United States HTA knowledge base. Rather than treating the United States as a single entity, each state is examined independently as a distinct academic and research environment. This approach recognizes that major states have developed substantial university-based programs in health economics, HTA and outcomes research that collectively shape graduate education, methodological research and healthcare decision-making. Each state interrogation applies the same twenty-four canonical statements derived from representational measurement, generating categorical probabilities and normalized logits that characterize the extent to which lawful measurement or measurement inversion defines the knowledge base.

California is an appropriate starting point. It contains one of the largest and most influential concentrations of HTA, health economics and outcomes research expertise in the United States. Its universities and research centers have made important contributions to comparative effectiveness research, health economics, patient-reported outcomes, economic evaluation and decision modelling. If representational measurement had become established anywhere as the scientific foundation for HTA, California would be expected to demonstrate that transition.

The interrogation reaches a different conclusion. California has faithfully inherited and reproduced the international HTA framework that has dominated the discipline for more than thirty years. The defining role of ratio measurement in supporting arithmetic has effectively disappeared from the knowledge base. In its place stands an analytical framework founded upon utilities, QALYs and modelled cost-effectiveness claims whose quantitative legitimacy depends upon ignoring the constraints imposed by representational measurement. The consequence is not simply methodological weakness but measurement inversion and, ultimately, measurement failure.

The objective of this paper is therefore broader than evaluating one state's educational and research environment. It asks whether the California HTA knowledge base is organized around the scientific principles governing quantitative measurement or whether it continues to reproduce a legacy framework that cannot satisfy the standards of normal science. The findings also point to a way forward. Once representational measurement is restored as the organizing scientific principle, HTA acquires a coherent and scientifically defensible analytical framework founded upon linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes. The California interrogation is the first step in documenting that transition across the United States.

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.

CALIFORNIA HEALTH TECHNOLOGY ASSESSMENT KNOWLEDGE BASE

California possesses one of the largest and most influential concentrations of health technology assessment (HTA), health economics and outcomes research expertise in the United States. Its universities, medical schools, schools of public health, pharmacy programs and affiliated research institutes undertake extensive research in comparative effectiveness, economic evaluation, health services research, patient-reported outcomes, decision science, implementation research, real-world evidence and healthcare policy. Collectively these activities constitute a substantial HTA knowledge base that influences graduate education, methodological development, healthcare policy and reimbursement decisions both nationally and internationally.

The California knowledge base extends well beyond formal HTA units. It encompasses university programs in health economics, pharmacoeconomics, comparative effectiveness research, outcomes research, value assessment, population health, health policy and quantitative evaluation. Major university systems, including the University of California campuses, Stanford University, the University of Southern California and associated medical centers, contribute through graduate teaching, multidisciplinary research programs, methodological publications, seminars, workshops and collaborations with government agencies, healthcare systems and the pharmaceutical and medical device industries. Together these institutions have established California as one of the principal centers for HTA and health economics research in North America.

The methodological orientation of this knowledge base is equally clear. Publicly available descriptions of research priorities, graduate teaching, faculty expertise and research outputs demonstrate comprehensive engagement with the contemporary analytical framework of health technology assessment. Economic evaluation, cost-effectiveness analysis, cost-utility analysis, patient-reported outcomes, health utility measurement, preference elicitation, decision analysis, simulation modelling, Markov modelling, microsimulation, real-world evidence, comparative effectiveness research and healthcare resource allocation are extensively represented. Sophisticated statistical analysis, predictive modelling and decision-analytic techniques are regarded as central components of graduate education and research. California therefore represents an advanced and technically accomplished HTA environment whose quantitative capabilities are widely recognized.

The present review, however, does not evaluate the technical sophistication, productivity or policy influence of these programs. 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 California HTA knowledge base explicitly recognize representational measurement as the foundation for quantitative claims before introducing the analytical methods of contemporary HTA? Specifically, does the 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 arithmetic operations performed within contemporary HTA satisfy the 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 those operations possess the measurement properties required by representational measurement, increasingly sophisticated statistical methods cannot establish their scientific legitimacy. The critical issue is therefore not whether California teaches contemporary HTA comprehensively—it clearly does—but whether the scientific conditions governing that quantitative framework are themselves explicitly represented.

California provides an especially demanding test of this proposition. If representational measurement had become established anywhere as the organizing scientific framework for HTA, it would reasonably be expected to appear within one of the most mature and methodologically sophisticated academic environments in the United States. The interrogation reported here therefore examines the California HTA 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 California's HTA enterprise or whether, as previous national and international investigations suggest, the knowledge base continues to reproduce the thirty-year 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

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

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 REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

The California interrogation identifies far more than isolated deficiencies in the teaching or practice of health technology assessment (Table 1). It demonstrates the successful transmission of a thirty-year international legacy of measurement inversion. California has inherited, reproduced and institutionalized the contemporary HTA framework without questioning its scientific foundations. The endorsement profile is internally consistent. Statements expressing the principles of lawful measurement receive uniformly weak endorsement, while false propositions that sustain utilities, QALYs, cost-effectiveness analysis and reference-case modelling receive uniformly strong endorsement. This is not an accidental pattern. It is the signature of a mature knowledge base that has learned to perform increasingly sophisticated arithmetic while progressively losing sight of the scientific conditions that make arithmetic possible.

The single most important finding concerns ratio measurement. The interrogation demonstrates that the defining role of ratio scales in supporting multiplication, division and ratio calculations has effectively disappeared from the California HTA knowledge base. The proposition that multiplication requires ratio measurement receives an endorsement probability of only 0.10 (−2.00 logits). Measurement preceding arithmetic receives the same endorsement 0.10 (−2.00). The axioms of representational measurement likewise receive 0.10 (−2.00). Together these statements define the scientific rules governing every quantitative discipline. Their weak endorsement demonstrates that the California knowledge base no longer recognizes the conditions that determine whether arithmetic is scientifically admissible. This omission is not a minor curricular weakness. It is the point at which the contemporary HTA framework collapses. Every central analytical construct depends upon arithmetic. Utilities are multiplied by time to generate QALYs. Incremental costs are divided by incremental QALYs to create cost-effectiveness ratios. Preference scores are aggregated, transformed and incorporated into increasingly elaborate simulation models. Every one of these operations presupposes lawful measurement. Once ratio measurement is restored to its proper scientific position, these constructs must first demonstrate that the quantities entering the arithmetic possess ratio properties. The California endorsement profile indicates that this question is simply never asked.

The reverse side of the endorsement profile is equally revealing. The false proposition that ratio measures may possess negative values receives an endorsement probability of 0.90 (+2.20 logits). The false proposition that EQ-5D-3L preference algorithms create interval measures also receives 0.90 (+2.20). The false proposition that the QALY is a ratio measure receives 0.90 (+2.20), while the false proposition that QALYs can be aggregated receives the maximum positive endorsement of 0.95 (+2.50). These are not peripheral misunderstandings. They are the propositions required to sustain contemporary cost-utility analysis. The profile shows that the California knowledge base strongly reinforces the very assumptions that lawful measurement rejects.

The treatment of subjective scores provides further evidence of the same inversion. The false proposition that summations of subjective instrument responses are ratio measures receives 0.90 (+2.20), as does the proposition that summing Likert-item scores creates a ratio measure. Numerical addition is therefore treated as though it can manufacture measurement. Yet adding ordered category scores does not establish equal units, a non-arbitrary zero, invariance or unidimensionality. Arithmetic can produce totals, but it cannot transform ordinal observations into lawful quantities. The strong endorsement of these statements demonstrates that the distinction between scoring and measurement has largely disappeared from the knowledge base.

The Rasch-related findings expose the corresponding absence of a defensible framework for latent attributes. The proposition that there are only two admissible classes of therapy-impact measurement the linear ratio and Rasch logit ratio receive 0.05 (−2.50). The proposition that subjective responses can be transformed into interval measurement only through Rasch rules also receives 0.05 (−2.50). Rasch logit ratio measurement as the required basis for evaluating latent therapy impact receives 0.05 (−2.50), while recognition that the relevant outcome is possession of the latent attribute receives only 0.15 (−1.75). The California knowledge base therefore provides virtually no recognition of the measurement framework required for attributes such as pain, fatigue, anxiety, physical function or need fulfilment.

Occasional applications of Rasch modelling do not alter this conclusion. Rasch analysis may appear in instrument development, item reduction or psychometric validation, but this does not establish Rasch measurement as the organizing scientific framework for HTA. The decisive question is whether latent therapy-impact claims are required to rest upon a unidimensional, invariant person–item continuum before arithmetic or statistical analysis proceeds. The endorsement profile indicates that they are not. Questionnaire scores remain treated as outcomes even when possession of the defined attribute has never been demonstrated.

The findings concerning falsifiability reveal another internal contradiction. The true proposition that non-falsifiable claims should be rejected receives comparatively positive endorsement at 0.75 (+1.10). Yet the false proposition that reference-case simulations generate falsifiable claims receives 0.85 (+1.75). The knowledge base therefore acknowledges falsifiability in principle while exempting the dominant modelling framework from its implications. A lifetime simulation is conditional upon assumptions, parameter selections, model structure and the chosen time horizon. Its output is not an empirical observation and cannot itself be directly replicated or falsified as a claim of therapy impact.

More importantly, modelling cannot repair measurement failure in its inputs. A simulation may be mathematically correct and internally coherent while remaining scientifically meaningless if the quantities being manipulated are not lawful measures. Utilities can be entered accurately, QALYs calculated precisely and ICERs reproduced exactly, yet none of these operations retrospectively establishes ratio measurement. Mathematics determines whether a calculation has been performed correctly. Measurement theory determines whether the result possesses the meaning claimed for it.

The endorsement profile therefore reveals two coherent but irreconcilable knowledge structures. The first is representational measurement: measurement preceding arithmetic, unidimensionality, ratio measurement, admissible transformations, dimensional homogeneity, linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes. These propositions receive uniformly weak endorsement. The second is contemporary HTA: utilities, QALYs, preference algorithms, summed scores, aggregation, cost-effectiveness ratios and reference-case modelling. The false propositions required to sustain this framework receive uniformly strong endorsement.

There is no halfway position between them. If representational measurement is accepted, its standards must be applied to every quantitative construct. Utilities must demonstrate ratio properties before multiplication by time; QALYs must satisfy dimensional homogeneity before aggregation; ICER denominators must represent lawful quantities before ratio formation; and latent attributes must be measured through Rasch requirements before claims of therapy impact are made. Contemporary HTA survives only while these questions remain outside the analytical framework.

The California interrogation therefore does not identify a collection of remediable methodological weaknesses. It demonstrates that California has inherited and institutionalized a thirty-year legacy of measurement inversion. Its universities and research centers may be highly accomplished in econometrics, evidence synthesis, statistical modelling, patient-reported outcomes and decision analysis, but technical sophistication cannot compensate for the disappearance of ratio measurement from the foundations of quantitative analysis. The omission is fatal because every principal HTA construct depends upon arithmetic whose scientific admissibility has not been established.

The implications are direct. What is presently taught and applied as quantitative HTA cannot generate scientifically defensible claims concerning therapy impact. Utilities, QALYs, cost-per-QALY ratios and reference-case simulations are not imperfect measures awaiting refinement. They are numerical constructions that fail the requirements governing lawful measurement. Once ratio measurement is restored as the non-negotiable foundation for multiplication, division and ratio formation, the contemporary framework collapses.

The California findings also contribute to a broader program of reproducible state-level interrogations. As the same 24 diagnostic statements are applied to other high-value state knowledge bases, the question will be whether California is exceptional or whether it reproduces a national structure of measurement inversion. If Texas, New York, Massachusetts, Pennsylvania and other major HTA environments produce comparable probability and logit profiles, the cumulative evidence will establish that the failure is disciplinary rather than local. The burden of proof will then shift decisively: the question will no longer be whether measurement inversion characterizes American HTA, but whether any major state knowledge base can demonstrate that it has escaped it.

California now has an opportunity to abandon this inherited legacy. The answer is not another refinement of utility estimation, another mapping algorithm or a more elaborate reference-case model. It is a new scientific beginning in which representational measurement becomes the only admissible analytical framework for HTA. This resolves the demarcation problem by requiring credible, empirically evaluable, replicable and falsifiable claims founded upon lawful measures.

The resulting framework is both simpler and more rigorous. Manifest therapy-impact attributes require linear ratio measures. Latent therapy-impact attributes require Rasch logit ratio measures**.** Arithmetic follows measurement; it does not create it. These two measures provide the only scientifically meaningful foundation for claims concerning therapy impact. California therefore has the opportunity not merely to correct an omission, but to establish a new standard for HTA by replacing thirty years of measurement inversion with the principles required of normal quantitative science.

IRRECONCILABLE PROGRAM CONTENT

A companion assessment of the California HTA program-content knowledge base extends the present findings beyond measurement inversion to a direct examination of what is actually taught and represented within graduate education and associated research programs [5]. The results are unequivocal. The scientific foundations of representational measurement are effectively absent.

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 framework is comprehensively represented through utilities, QALYs, cost-effectiveness analysis, preference elicitation, decision analysis, simulation modelling and reference-case methods.

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 California program-content review therefore demonstrates more than an educational omission. It identifies a curriculum structured around analytical methods that depend for their continued acceptance upon the absence of representational measurement. The omission is not incidental it is the condition that allows the contemporary HTA framework to survive.

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 [6]. 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 California interrogation demonstrates the successful institutionalization of a thirty-year legacy of 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 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 ratio measures, the analytical structure of contemporary HTA can no longer be sustained. Increasingly sophisticated modelling cannot compensate for the absence of lawful measurement.

The companion program-content assessment reaches 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 California HTA curriculum, while the methods of contemporary HTA are comprehensively represented. This is not 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 implications are unavoidable. Contemporary HTA, as presently constituted, should no longer be taught as a graduate quantitative discipline. Universities have a responsibility to ensure that quantitative subjects satisfy the accepted standards governing measurement before teaching the arithmetic that depends upon those standards. California now has an opportunity to abandon a thirty-year legacy of measurement inversion and establish a new scientific foundation for HTA. Representational measurement resolves the demarcation problem by restoring measurement before arithmetic and limiting claims of therapy impact to the only two admissible measures: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. This is not another methodological option within contemporary HTA. It is the only analytical framework consistent with the standards of normal quantitative science.

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] California: Normal science and the absence of representational measurement in HTA education: Logit Working Paper No 1855. https://maimonresearch.com/logit-working-paper-no-1855-july-2026/

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