MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

FLORIDA: 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 1905 AUGUST 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF MEASUREMENT INVERSION

ABSTRACT

Health technology assessment (HTA) is widely presented as a quantitative scientific discipline supporting claims concerning the comparative value of pharmaceutical products, medical devices and healthcare interventions. These claims increasingly influence reimbursement, pricing, formulary management and healthcare resource allocation. Their scientific legitimacy, however, depends upon a prior requirement: that the quantities entering economic evaluation satisfy the accepted standards of representational measurement. Measurement must precede arithmetic. Unless the quantities subjected to multiplication, division and ratio formation possess lawful measurement properties, the resulting claims cannot be regarded as scientifically meaningful. This paper examines whether the Florida HTA knowledge base recognizes these foundational scientific principles or instead reproduces the measurement inversion that characterizes contemporary HTA internationally.

The analysis forms part of a series of state-level interrogations of the United States HTA knowledge base. Florida's universities, colleges of pharmacy, schools of public health, medical schools and associated research centers collectively represent a substantial academic environment for health economics, outcomes research and health technology assessment. The investigation applies twenty-four canonical statements derived from representational measurement theory and Rasch measurement to this knowledge base, generating categorical endorsement probabilities and normalized logits that characterize the extent to which lawful measurement or measurement inversion defines the state's HTA framework. A companion program-content review examines whether representational measurement is explicitly incorporated within graduate teaching and research programs before students are introduced to conventional HTA methodologies.

The interrogation demonstrates a consistent pattern of measurement inversion. Fundamental scientific propositions concerning measurement preceding arithmetic, ratio measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, the distinction between manifest and latent attributes and Rasch measurement receive uniformly weak endorsement, whereas propositions supporting utilities, QALYs, cost-effectiveness analysis and reference-case modelling receive consistently strong endorsement. The companion program-content assessment reaches the same conclusion. Representational measurement is effectively absent from the Florida curriculum while the analytical methods of conventional HTA are comprehensively represented. Together these findings indicate that the scientific principles governing lawful quantitative measurement have been displaced by an inherited technical framework whose arithmetic is *undertaken without first establishing that the quantities involved possess the required measurement properties.*

The paper concludes that the Florida HTA knowledge base reflects the same international pattern observed previously in Australia, New Zealand, Canada, California, Texas and other jurisdictions. Contemporary HTA is reproduced as technical instruction rather than education in quantitative science. A transition to representational measurement is therefore required, placing measurement before arithmetic and restricting claims of therapy impact to lawful linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes. This provides a coherent scientific foundation for HTA based upon empirically evaluable and potentially falsifiable claims rather than numerical constructions unsupported by lawful measurement.

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 forms part of 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 many 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.

Florida provides an important setting for this investigation**.** The state has developed a substantial network of universities, colleges of pharmacy, schools of public health, medical schools and health services research centers that contribute to education and research in health economics, outcomes research and health technology assessment. Collectively, these institutions influence graduate education, methodological development and healthcare decision-making throughout Florida. If representational measurement had become established as the scientific foundation for HTA within this educational environment, the Florida knowledge base would be expected to demonstrate that transition.

The interrogation reaches a different conclusion. Florida 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 Florida 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 Florida interrogation represents a further step in documenting that transition across the United States.

I actually think this paragraph is **stronger** than the California version. California's claim rested on its size and influence. Florida's strength lies in showing that even without California's concentration of internationally recognized HTA centers, exactly the same pattern of measurement inversion emerges. That reinforces your broader thesis that measurement inversion is a property of the international HTA paradigm itself rather than of particular universities or 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.

THE FLORIDA HTA KNOWLEDGE BASE

Florida has developed a substantial and geographically diverse health technology assessment (HTA), health economics and outcomes research knowledge base distributed across its major public and private universities. Colleges of pharmacy, schools of public health, medical schools, health services research institutes and interdisciplinary research centers collectively undertake research in comparative effectiveness, economic evaluation, pharmacoeconomics, outcomes research, implementation science, patient-reported outcomes, population health, healthcare policy and healthcare delivery. Although Florida has no single institution dedicated exclusively to HTA, the combined activities of these academic units constitute an extensive educational and research environment that contributes to graduate education, methodological development and healthcare decision-making throughout the state.

The Florida HTA knowledge base extends well beyond formal programs in health economics. It encompasses graduate teaching and research in pharmacoeconomics, comparative effectiveness research, health services research, value assessment, population health, epidemiology, healthcare management, pharmaceutical outcomes research and evidence-based medicine. Universities including the University of Florida, the University of South Florida, Florida State University, the University of Miami, Florida International University, the University of Central Florida, Florida Atlantic University, Nova Southeastern University, the University of North Florida and Florida A&M University contribute through multidisciplinary research programs, graduate curricula, seminars, methodological publications and collaborations with healthcare systems, government agencies, professional organizations and the pharmaceutical and medical device industries. Together these institutions provide Florida with one of the largest university-based health outcomes research environments in the southeastern United States.

The methodological orientation of this knowledge base is readily apparent from publicly available descriptions of research priorities, faculty expertise, graduate teaching and research outputs. Economic evaluation, cost-effectiveness analysis, cost-utility analysis, health utility measurement, patient-reported outcomes, preference elicitation, comparative effectiveness research, decision analysis, simulation modelling, real-world evidence, healthcare utilization, value assessment and health policy analysis are extensively represented. Sophisticated statistical methods, predictive analytics, decision-analytic modelling and economic evaluation are presented as core components of graduate education and research. Florida therefore possesses a technically sophisticated HTA environment whose research contributes to both regional and national discussions of healthcare value and resource allocation.

The present review does not evaluate the productivity, policy influence or technical competence of these programs. Nor does it assess the scientific contributions of individual investigators. The question addressed is considerably narrower but scientifically more fundamental. Does the Florida HTA knowledge base explicitly recognize representational measurement as the scientific foundation for quantitative claims before introducing the analytical methods of contemporary

HTA? Specifically, does it 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 performed within contemporary HTA satisfies the accepted standards 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 analytical methods cannot establish their scientific legitimacy. The interrogation reported here therefore examines the Florida HTA knowledge base using twenty-four canonical statements derived from representational measurement. The objective is to determine whether lawful measurement provides the scientific foundation for Florida's HTA enterprise or whether, consistent with findings from other jurisdictions, the knowledge base continues to reproduce the international pattern of measurement inversion that has characterized contemporary health technology assessment for more than three decades.

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 FLORIDA

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 interrogation of the Florida HTA knowledge base demonstrates a consistent pattern of measurement inversion. Of the twenty-four canonical propositions, only one of the scientific propositions required to support lawful quantitative measurement receives strong endorsement. Time is a ratio measure is endorsed with a probability of 0.95 (normalized logit +2.50), reflecting a universally accepted principle of physical measurement. Beyond this, however, the pattern changes dramatically. The propositions that define representational measurement receive uniformly weak endorsement. Interval measures lack a true zero is endorsed at only 0.20 (logit

−1.40), measures must be unidimensional at 0.15 (−1.75), multiplication requires a ratio measure at 0.10 (−2.00), measurement precedes arithmetic at 0.10 (−2.00), and meeting the axioms of representational measurement is required for arithmetic also at 0.10 (−2.00). Even more striking is the almost complete absence of recognition for the propositions that distinguish lawful measurement of latent attributes. There are only two classes of measurement—linear ratio and Rasch logit ratio, transforming subjective responses to interval measurement is only possible with Rasch rules, and the Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits each receive endorsement probabilities of only 0.05, corresponding to the minimum normalized logit of −2.50. Likewise, the Rasch rules for measurement are identical to the axioms of representational measurement is endorsed at 0.05 (−2.50), while the outcome of interest for latent traits is the possession of that trait receives only 0.15 (−1.75).

In direct contrast, the knowledge base provides consistently strong endorsement for propositions supporting the conventional analytical framework of contemporary HTA. The false propositions time trade-off preferences are unidimensional, ratio measures can have negative values, EQ-5D-3L preference algorithms create interval measures, the QALY is a ratio measure, summations of subjective instrument responses are ratio measures, summation of Likert question scores creates a ratio measure, and the QALY is a dimensionally homogeneous measure each receive endorsement probabilities of 0.90 (normalized logits +2.20). Similarly, QALYs can be aggregated receives an endorsement probability of 0.95 (+2.50), while reference-case simulations generate falsifiable claims is endorsed at 0.85 (+1.75). These probabilities indicate that propositions inconsistent with representational measurement are treated as established components of the Florida HTA knowledge base.

The remaining items reinforce this pattern. Although non-falsifiable claims should be rejected and the logit is the natural logarithm of the odds ratio each receive relatively strong endorsement (0.75, logit +1.10), they are isolated scientific propositions embedded within a broader analytical framework that fails to recognize the measurement conditions necessary for their meaningful application. Similarly, a linear ratio scale for manifest claims can always be combined with a logit scale receives only modest rejection (0.60, logit +0.40), indicating uncertainty regarding the distinction between manifest and latent measurement systems.

Taken together, these results demonstrate that the Florida HTA knowledge base is organized around measurement inversion rather than representational measurement. The scientific propositions that should define the foundations of quantitative inquiry consistently receive endorsement probabilities between 0.05 and 0.20, with normalized logits ranging from −2.50 to −1.40, while propositions supporting utilities, QALYs, preference algorithms, cost-effectiveness analysis and reference-case modelling receive endorsement probabilities between 0.85 and 0.95, with normalized logits between +1.75 and +2.50. The polarity of the logit distribution is therefore unmistakable: lawful measurement is weakly endorsed while the conventional HTA analytical framework is strongly endorsed.

This profile is essentially identical to that observed in Australia, California, Texas, Canada, New Zealand, the United Kingdom and other HTA jurisdictions examined to date. The Florida findings should therefore not be interpreted as reflecting the practices of individual universities or research centers. Rather, they demonstrate that Florida participates in the same international HTA paradigm in which arithmetic is assumed to confer quantitative meaning without first establishing that the quantities involved satisfy the accepted standards of representational measurement. The Florida HTA knowledge base consequently reproduces measurement inversion as the organizing scientific principle of education, research and professional practice.

IRRECONCILABLE PROGRAM CONTENT

A companion assessment of the Florida 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. 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 Florida 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 [5]. Designed specifically for HTA, the program provides a structured transition from the legacy framework represented by the ISPOR Good Practices recommendations to one grounded in representational measurement. Beginning with the principle that measurement precedes arithmetic, the nine units progressively address the definition of attributes, scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession and the construction of empirically evaluable therapy-impact claims.

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

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

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

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

CONCLUSION

The Florida 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 Florida 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. Florida's universities now have 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] Maimon Research. HTA Reconstruction. https://maimonresearch.com/hta-reconstruction-program-and-fees/