

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**FRANCE: THE ENDORSEMENT OF MEASUREMENT
INVERSION IN THE HAUTE AUTORITÉ DE SANTÉ
(HAS)**

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

LOGIT WORKING PAPER No 1475 JULY 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF NON-MEASUREMENT

ABSTRACT

This Logit Working Paper examines the scientific foundations of the French health technology assessment (HTA) framework through an interrogation of the publicly accessible knowledge base of the Haute Autorité de Santé (HAS). The objective is not to evaluate individual reimbursement decisions or administrative procedures but to determine whether the methodological framework promoted by HAS recognizes the principles of representational measurement that are required before lawful quantitative claims regarding therapy impact can be advanced. The interrogation applies a fixed set of twenty-four canonical statements derived from the axioms of representational measurement, the theory of scale construction, Rasch measurement, dimensional homogeneity, and the philosophy of falsifiable scientific inquiry.

The interrogation demonstrates a consistent pattern of measurement inversion. Statements representing the scientific foundations of quantitative measurement including the requirement that measurement precede arithmetic, the necessity for ratio scales to support multiplication, the role of unidimensionality, the distinction between manifest and latent attributes, and the central importance of Rasch measurement for latent constructs receive uniformly weak endorsement. In contrast, propositions underpinning contemporary HTA practice, including utilities, EQ-5D preference algorithms, QALYs, cost-effectiveness ratios, aggregation of QALYs, and reference-case simulation modelling, receive consistently strong endorsement despite their incompatibility with the requirements of representational measurement.

The findings indicate that the HAS methodological framework gives priority to accepted analytical techniques while largely overlooking the scientific conditions necessary to justify those techniques. Arithmetic is permitted to precede measurement, numerical convention substitutes for validated measurement, and simulation modelling is accepted without first establishing the measurement properties of the quantities being manipulated. The result is a framework that is administratively sophisticated yet detached from the measurement principles that govern quantitative reasoning throughout the physical and social sciences.

The paper argues that these findings have implications extending well beyond HAS itself. As the principal HTA authority in France, HAS defines the methodological expectations that influence academic research, professional education, consultancy practice, reimbursement submissions, and the wider French HTA community. Measurement inversion therefore becomes institutionalized and is subsequently reproduced through education and professional practice. The paper concludes that incremental refinement of reference-case methodology cannot resolve these deficiencies. Reconstruction must begin with the restoration of representational measurement as the scientific foundation of HTA, employing linear ratio measures for manifest attributes, Rasch

logit ratio measures for latent attributes, and prospectively specified, evaluable, replicable, and falsifiable claims regarding therapy impact. In this framework, HTA becomes a continuing process of empirical scientific inquiry rather than an administrative exercise designed primarily to achieve reimbursement closure.

INTRODUCTION

The Haute Autorité de Santé (HAS) is the principal organization responsible for health technology assessment (HTA) in France and occupies a central position in the country's healthcare decision-making process. Established as an independent public authority, HAS develops methodological guidance, evaluates pharmaceuticals, medical devices, diagnostic technologies, clinical procedures, and public health interventions, and provides recommendations that inform reimbursement, pricing negotiations, clinical practice, and resource allocation within the French health system. Through its specialist committees—including the Transparency Committee, the Economic and Public Health Evaluation Committee (CEESP), and the National Committee for the Evaluation of Medical Devices and Health Technologies (CNEDiMTS)—HAS has developed a comprehensive methodological framework for assessing clinical effectiveness, comparative benefit, safety, economic value, and organizational impact.

Beyond its regulatory responsibilities, HAS exerts a major influence on the evolution of HTA in France through the publication of methodological guidance and technical recommendations that shape the analytical standards expected of manufacturers, researchers, consultants, academic institutions, and professional organizations. Consequently, the publicly accessible HAS knowledge base represents the authoritative statement of the methodological principles that underpin contemporary French HTA and provides an appropriate basis for evaluating the extent to which the scientific requirements of representational measurement are recognized within the national HTA framework.

The objective of the present study is to evaluate the extent to which the publicly accessible knowledge base of the Haute Autorité de Santé (HAS) recognizes the scientific principles required to support lawful quantitative claims regarding therapy impact. The interrogation applies a fixed set of twenty-four canonical statements derived from the axioms of representational measurement and the requirements for both manifest and latent attribute measurement. These statements encompass the principal scales of measurement, ratio measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, the distinction between manifest and latent attributes, Rasch measurement, falsifiability, and the scientific legitimacy of utilities, QALYs, cost-effectiveness ratios, and reference-case simulation modelling. The purpose is not to evaluate individual HTA reports or committee decisions but to determine whether the methodological framework communicated through the HAS knowledge base provides the scientific foundations necessary for quantitative therapy assessment or whether it instead embodies the characteristics of measurement inversion.

The objective of the present study is to evaluate the extent to which the publicly accessible knowledge base of the Haute Autorité de Santé (HAS) recognizes the scientific principles required to support lawful quantitative claims regarding therapy impact. The interrogation applies a fixed set of twenty-four canonical statements derived from the axioms of representational measurement

and the requirements for both manifest and latent attribute measurement. These statements encompass the principal scales of measurement, ratio measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, the distinction between manifest and latent attributes, Rasch measurement, falsifiability, and the scientific legitimacy of utilities, QALYs, cost-effectiveness ratios, and reference-case simulation modelling. The purpose is not to evaluate individual HTA reports or committee decisions but to determine whether the methodological framework communicated through the HAS knowledge base provides the scientific foundations necessary for quantitative therapy assessment or whether it instead embodies the characteristics of measurement inversion.

The interrogation demonstrates a consistent pattern of measurement inversion throughout the HAS knowledge base. Statements representing the scientific foundations of measurement receive uniformly weak endorsement, while propositions supporting utilities, EQ-5D preference algorithms, QALYs, cost-effectiveness ratios, aggregation of QALYs, and reference-case modelling receive consistently strong endorsement despite their incompatibility with the axioms of representational measurement. The interrogation indicates that the HAS methodological framework gives priority to established analytical techniques while providing comparatively little recognition to the measurement principles that must precede quantitative analysis. Arithmetic is therefore permitted to precede measurement, rather than being constrained by it. The implication is that the French national HTA framework rests upon methodological conventions that have not first demonstrated the measurement properties necessary to support lawful quantitative claims. This conclusion provides the basis for the companion assessment of curriculum inversion, where the question becomes whether these same methodological assumptions are reproduced through the educational and professional training environment that prepares future HTA practitioners.

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 scales¹. 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 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

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 HAS HTA KNOWLEDGE BASE

The first step in any large language model (LLM) interrogation is to define the knowledge base to be examined. The validity of the interrogation depends upon ensuring that the knowledge base accurately represents the intellectual environment within which methodological concepts are communicated, interpreted, and reinforced. A knowledge base is considerably more than a collection of documents. It comprises the accumulated concepts, assumptions, analytical methods, definitions, guidance, and decision frameworks through which an institution establishes and communicates its understanding of a discipline. In the context of health technology assessment (HTA), the knowledge base provides the conceptual framework that informs manufacturers, academic researchers, consultants, committee members, and decision makers regarding acceptable standards of evidence and quantitative evaluation.

For the present interrogation, the HAS HTA knowledge base is defined as the totality of publicly accessible materials through which HAS communicates its methodological framework for health technology assessment. These materials include methodological guidance documents, health economic evaluation manuals, CEESP methodological recommendations, Transparency Committee guidance, CNEDiMTS documentation for medical devices and health technologies, scientific opinions, technical reports, consultation papers, explanatory publications, implementation guidance, methodological updates, workshop and seminar materials, public presentations, policy documents, and associated publications available through the HAS website and related public resources. Collectively, these materials define the analytical framework that applicants, reviewers, researchers, consultants, and policy makers are expected to understand and apply when preparing, evaluating, and interpreting HTA submissions within France.

The importance of defining the HAS knowledge base extends beyond the content of individual guidance documents. National HTA agencies do more than evaluate evidence. They establish methodological norms. Through their published guidance they determine which analytical techniques are regarded as scientifically acceptable, which forms of evidence are considered legitimate, which outcome measures are appropriate for reimbursement decisions, and which economic methods are expected in manufacturer submissions. These methodological expectations subsequently influence university teaching, professional training, consultancy practice, journal publication standards, and the wider research community. In this way, the HAS knowledge base functions not merely as an administrative resource but as a principal mechanism through which methodological assumptions become institutionalized within French HTA.

A national HTA knowledge base designed to support lawful quantitative claims would be expected to recognize the scientific principles that govern measurement before numerical analysis is undertaken. These principles include specification of the target attribute, the principal scales of measurement, the axioms of representational measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, the distinction between manifest and latent attributes, the requirement for linear ratio measures for manifest attributes and Rasch logit ratio measures for

latent attributes, together with the requirement that all claims regarding therapy impact should be prospectively evaluable, replicable, and capable of potential falsification. These are not optional methodological refinements. They constitute the scientific foundations upon which all quantitative disciplines depend.

The objective of the present interrogation is therefore not to evaluate the administrative performance of HAS, the quality of individual assessments, or the competence of its scientific committees. Rather, it is to examine the broader methodological environment represented by the HAS knowledge base and determine the extent to which the concepts required for measurement-based HTA are explicitly recognized. The central question is straightforward: does the publicly accessible methodological framework of HAS expose users to the scientific principles required to support lawful quantitative claims regarding therapy impact, or does it primarily promote established analytical techniques without first establishing their measurement foundations? The answer provides an indication of whether the French national HTA framework is grounded in representational measurement or whether it embodies the characteristic features of measurement inversion.

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 endorsement probability: the estimated likelihood that the HTA knowledge base endorses the statement whether it is true or false; *explicitly or implicitly*.

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

This is why categorical probabilities are essential. Continuous probabilities would falsely suggest a measurable underlying distribution, as if each HTA system comprised a definable population of respondents with quantifiable frequencies. But large language models do not operate on that level. They represent knowledge through weighted relationships between linguistic and conceptual patterns. When asked whether a domain tends to affirm, deny, or ignore a principle such as

unidimensionality, admissible arithmetic, or the axioms of representational measurement, the model draws on its internal structure to produce an estimate of conceptual reinforcement. The precision of that estimate must match the nature of the task. Categorical probabilities therefore provide a disciplined and interpretable way of capturing reinforcement strength while avoiding the illusion of statistical granularity.

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

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

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

The LLM computes these categorical probabilities from three sources:

1. **Structural content of HTA discourse**

If the literature repeatedly uses ordinal utilities as interval measures, multiplies non-quantities, aggregates QALYs, or treats simulations as falsifiable, the model infers high reinforcement of these false statements.

2. **Conceptual visibility of measurement axioms**

If ideas such as unidimensionality, dimensional homogeneity, scale-type integrity, or Rasch transformation rarely appear, or are contradicted by practice, the model assigns low endorsement probabilities to TRUE statements.

3. **The model's learned representation of domain stability**

Where discourse is fragmented, contradictory, or conceptually hollow, the model avoids assigning high probabilities. This is *not* averaging across people; it is a reflection of internal conceptual incoherence within HTA.

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

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

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

INTERROGATION STATEMENTS

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

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

Measurement Theory & Scale Properties

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

Measurement Preconditions for Arithmetic

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

Rasch Measurement & Latent Traits

12. There are only two classes of measurement: linear ratio and Rasch logit ratio — TRUE

13. Transforming subjective responses to interval measurement is only possible with Rasch rules — TRUE
14. Summation of Likert question scores creates a ratio measure — FALSE

Properties of QALYs & Utilities

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

Falsifiability & Scientific Standards

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

Logit Fundamentals

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

Latent Trait Theory

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

AI LARGE LANGUAGE MODEL STATEMENTS: TRUE OR FALSE

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

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

INTERPRETING TRUE STATEMENTS

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

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

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

INTERPRETING FALSE STATEMENTS

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

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

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

2. SUMMARY OF FINDINGS FOR TRUE AND FALSE ENDORSEMENTS: THE HAS HTA KNOWLEDGE BASE

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

It is essential to understand how to interpret these results.

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

The logits provide a continuous, symmetric scale, ranging from +2.50 to -2.50, that quantifies the degree of this endorsement. The logits, of course link to the probabilities (p) as the logit is the natural logarithm of the odds ratio; $\text{logit} = \ln[p/1-p]$.

- 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.

THE HAS HTA KNOWLEDGE BASE: ENDORSING FALSE MEASUREMENT

The interrogation of the publicly accessible knowledge base of HAS provides a revealing picture of the conceptual foundations upon which contemporary French health technology assessment (HTA) has been constructed (Table 1). The knowledge base encompasses the methodological guidance, economic evaluation manuals, committee documentation, reimbursement guidance, technical reports, consultation papers, and explanatory documents through which HAS communicates the analytical framework expected of manufacturers, academic researchers, consultants, and decision makers. Collectively these materials define the intellectual environment within which quantitative claims regarding therapy impact are generated, interpreted, and accepted.

TABLE 1: ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS: THE HAS HTA KNOWLEDGE BASE

STATEMENT	RESPONSE 1=TRUE 0=FALSE	ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY	NORMALIZED LOGIT (IN RANGE +/- 2.50)
INTERVAL MEASURES LACK A TRUE ZERO	1	0.10	-2.20
MEASURES MUST BE UNIDIMENSIONAL	1	0.10	-2.20
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.05	-2.50
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	0.85	+1.70
THE QALY IS A RATIO MEASURE	0	0.90	+2.20
TIME IS A RATIO MEASURE	1	0.85	+1.70
MEASUREMENT PRECEDES ARITHMETIC	1	0.03	-2.50
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.85	+1.70
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.05	-2.50
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	0.85	+1.70
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.90	+2.20
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

QALYS CAN BE AGGREGATED	0	0.90	+2.20
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.15	-1.70
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.85	+1.70
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.05	-2.50
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	0.85	+1.70
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.05	-2.50
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

The interrogation does not assess the administrative effectiveness of HAS, the quality of its staff, or the sophistication of its procedural guidance. Those are separate questions. Rather, it asks a more fundamental scientific question: does the HAS knowledge base recognize the conditions required before quantitative claims regarding therapy impact can legitimately be advanced? More specifically, does it recognize the principles of representational measurement that govern the assignment of numbers to attributes and determine the admissibility of arithmetic operations? The twenty-four canonical statements address precisely these issues.

The endorsement profile is striking. Statements that represent the foundations of measurement science receive consistently weak endorsement, while statements that support utilities, QALYs, preference algorithms, and reference-case methodology receive consistently strong endorsement. This is precisely the profile expected where measurement inversion has become institutionalized. Arithmetic has assumed precedence over measurement, numerical convention has displaced scientific validation, and accepted methodology has become detached from the principles that govern quantitative reasoning.

The weakness begins with the concept of measurement itself. The interrogation indicates only limited recognition that interval measures lack a true zero and that multiplication requires ratio measurement. These are elementary propositions within representational measurement. The distinction between interval and ratio scales is not a matter of technical preference but determines the mathematical operations that may legitimately be performed. A ratio scale permits proportional comparison because it possesses a non-arbitrary zero representing the complete absence of the

attribute. Interval scales do not. If this distinction is ignored, multiplication and division become mathematically inadmissible regardless of the sophistication of subsequent statistical analysis.

This omission is particularly significant because contemporary HTA relies extensively upon arithmetic operations that implicitly assume ratio measurement. Costs are divided by outcome measures to produce incremental cost-effectiveness ratios. Utility values are multiplied by time to generate quality-adjusted life years. Lifetime simulations repeatedly aggregate these quantities over extended periods. Yet the interrogation indicates little recognition within the HAS knowledge base that such arithmetic requires ratio measurement as a prior condition. The scientific sequence has therefore been reversed. Rather than first establishing the measurement properties of the quantities being manipulated, arithmetic proceeds as though measurement had already been demonstrated.

The interrogation also indicates only weak endorsement of the proposition that measures must be unidimensional. This omission has profound implications. Measurement is possible only when observations refer to a single underlying attribute. If multiple attributes are combined within a composite score, aggregation does not create measurement. It merely creates another numerical summary. The requirement for unidimensionality is therefore not a refinement of psychometric theory but a prerequisite for quantitative science. Without it, there is no assurance that a numerical value corresponds to a single measurable property.

This issue becomes especially important when considering preference-based measures of health status. Contemporary HTA routinely employs instruments that combine mobility, self-care, usual activities, pain, anxiety, and other domains into a single preference-weighted index. Such instruments are undoubtedly useful descriptive summaries of health status, but their aggregation of multiple dimensions immediately raises the question of whether a single attribute has been measured. The interrogation suggests that this question receives little explicit attention within the HAS knowledge base. Instead, multidimensional preference scores are accepted as though they possessed the properties required for subsequent arithmetic.

Closely related is the limited endorsement of the proposition that measurement must precede arithmetic. This statement lies at the heart of representational measurement and, indeed, of all quantitative science. Before numerical operations can legitimately be performed, the numerical values themselves must first be shown to constitute measures. Scientific disciplines outside HTA rarely confuse these two stages. Physics does not multiply quantities before establishing the units involved. Chemistry does not manipulate concentrations before defining the relevant scales. Engineering does not undertake structural calculations before establishing the dimensions of the variables being analyzed. Yet the interrogation indicates that the HAS knowledge base gives comparatively little recognition to this elementary scientific principle.

The consequence is the emergence of an analytical culture in which accepted methodology substitutes for demonstrated measurement. Numerical sophistication becomes confused with scientific validity. Increasingly elaborate statistical models, probabilistic sensitivity analyses, Bayesian updating, and lifetime simulations create an appearance of quantitative precision while leaving unresolved the more fundamental question of whether the quantities entering those analyses possess the measurement properties that the arithmetic requires.

Equally revealing is the weak endorsement of the proposition that arithmetic itself depends upon the axioms of representational measurement. These axioms define the conditions under which empirical relationships may legitimately be represented numerically. They establish the logical connection between observable phenomena and mathematical operations. Without them there is no scientific justification for claiming that numerical manipulations correspond to empirical reality. Their absence from the HAS knowledge base therefore represents more than an academic omission. It removes the scientific foundation upon which quantitative HTA claims should rest.

The interrogation similarly reveals little recognition that only two forms of ratio measurement are available for HTA. Manifest attributes—those capable of direct observation—require linear ratio measures. Latent attributes—those that cannot be observed directly—require Rasch logit ratio measures. This distinction represents one of the most important developments in modern measurement theory. It recognizes that observable phenomena such as mortality, hospital admissions, treatment persistence, and resource utilization differ fundamentally from latent phenomena such as pain, fatigue, physical functioning, treatment satisfaction, need fulfilment, and quality of life. Each class of attribute requires its own measurement framework.

The limited endorsement of this proposition within the HAS knowledge base has important consequences. Without distinguishing manifest from latent attributes, all numerical outcomes tend to be treated as though they occupied the same measurement space. Observable counts, ordinal questionnaire responses, preference weights, composite indices, and statistical estimates become interchangeable inputs into economic models despite possessing fundamentally different measurement properties. The result is not simply conceptual confusion but the systematic erosion of measurement standards.

Particularly noteworthy is the near absence of recognition that transformation of subjective responses into interval measurement is only possible through Rasch measurement. This is not a statement about statistical convenience but about the conditions under which latent attributes may legitimately be measured. Rasch measurement was developed precisely because latent constructs cannot be observed directly. Rather than treating questionnaire responses as measures, Rasch models estimate possession of the underlying attribute while simultaneously testing unidimensionality, invariance, response functioning, and model fit. Without this framework, subjective responses remain ordinal observations regardless of the sophistication of subsequent statistical manipulation.

The HAS interrogation suggests that this distinction is largely absent from the publicly accessible methodological framework. Patient-reported outcomes and health-related quality of life are acknowledged as important components of HTA, yet the scientific conditions required for measuring these latent attributes receive little visible emphasis. Consequently, subjective responses are liable to be treated as though they were already quantitative measures, allowing arithmetic operations to proceed without the necessary measurement foundations.

The interrogation nevertheless identifies one area of strong endorsement. Time is correctly recognized as a ratio measure. This is entirely consistent with physical measurement. Time possesses a true zero, equal units, and supports all ratio operations. Its recognition, however, serves only to emphasize the inconsistency elsewhere within the HAS framework. Time satisfies the

conditions for ratio measurement; the utility values with which it is multiplied do not. The existence of one lawful ratio measure therefore cannot compensate for the absence of another. Multiplication remains inadmissible unless both operands satisfy the requirements of ratio measurement. The recognition of time as a ratio measure therefore highlights, rather than resolves, the measurement deficiencies associated with QALY construction.

The interrogation becomes even more revealing when attention turns to the propositions that contemporary HTA routinely accepts as methodological orthodoxy. Here the endorsement profile is effectively reversed. Statements that are inconsistent with the principles of representational measurement receive consistently strong endorsement, demonstrating that the HAS knowledge base has incorporated assumptions that measurement theory would reject. These statements concern time trade-off preferences, EQ-5D preference algorithms, utilities, QALYs, cost-effectiveness ratios, aggregation, reference-case simulation modelling, and the treatment of subjective questionnaire responses. Collectively they define the architecture of contemporary reference-case HTA.

The interrogation indicates strong endorsement of the proposition that time trade-off preferences may be regarded as unidimensional. This assumption is central to the construction of utility instruments because the preference values generated by time trade-off exercises become the numerical inputs to QALY calculations. Yet a preference exercise is not a measurement model. Respondents evaluate hypothetical health states that themselves comprise multiple dimensions of health, personal expectations, social context, adaptation, and risk attitudes. There is no demonstration that the resulting preferences represent possession of a single latent attribute. Rather, they reflect complex subjective judgements influenced by multiple considerations. To regard such preferences as unidimensional measures is therefore to bypass the very conditions that measurement requires.

The same conclusion applies to the strong endorsement of the proposition that EQ-5D preference algorithms create interval measures. Preference algorithms undoubtedly generate numerical values, but arithmetic transformation cannot manufacture measurement where the underlying observations remain ordinal. The application of tariffs, regression models, or preference weights alters numerical representation; it does not establish that the resulting values satisfy the axioms of representational measurement. The interrogation therefore suggests that the HAS knowledge base accepts the numerical outputs of EQ-5D algorithms as though the act of algorithmic transformation itself confers measurement status. This is a clear example of arithmetic substituting for measurement.

Perhaps the most important finding concerns endorsement of the proposition that the QALY is a ratio measure. This proposition lies at the centre of economic evaluation because the validity of cost-per-QALY ratios depends entirely upon the denominator possessing ratio-scale properties. The interrogation demonstrates strong endorsement of the QALY despite the simultaneous absence of recognition that multiplication requires ratio measurement, that ratio scales require true zero, and that negative values are incompatible with ratio measurement. Taken together these findings reveal a profound internal inconsistency. The knowledge base accepts the conclusion while giving little recognition to the scientific premises upon which that conclusion depends.

This inconsistency is reinforced by the endorsement of the proposition that ratio measures may possess negative values. Within representational measurement this is impossible. A true ratio scale requires a zero corresponding to complete absence of the attribute. Values below zero imply less than complete absence, an empirical impossibility. Yet preference-based utility systems employed throughout HTA routinely generate negative utilities for health states considered worse than death. Once such values are accepted, proportional comparison becomes meaningless, and multiplication by time cannot generate a lawful ratio measure. The interrogation therefore demonstrates that the HAS knowledge base accepts numerical conventions fundamentally incompatible with ratio measurement.

The strong endorsement of summed subjective questionnaire responses and summed Likert scores as ratio measures reinforces the same pattern. Summation is an arithmetic operation, not a measurement operation. Adding ordinal responses produces another ordinal summary unless a valid measurement model demonstrates otherwise. The widespread use of summed questionnaire scores within health outcomes research reflects convenience rather than measurement. The interrogation indicates that the HAS knowledge base provides little explicit distinction between descriptive scoring systems and scientifically validated measures. Consequently, arithmetic is permitted to substitute for measurement, perpetuating the very inversion identified throughout this programme of interrogations.

The endorsement of the proposition that the QALY is dimensionally homogeneous is equally revealing. Dimensional homogeneity requires that arithmetic operations combine quantities possessing compatible dimensions. Time is a physical ratio measure. Utilities derived from preference instruments are subjective numerical constructions whose measurement properties remain unestablished. Their multiplication therefore combines fundamentally different dimensions without demonstrating mathematical compatibility. The resulting QALY inherits this dimensional inconsistency. Yet the interrogation indicates that the HAS knowledge base accepts dimensional homogeneity as an implicit assumption of economic evaluation rather than as a proposition requiring scientific demonstration.

Closely related is the rejection of the proposition that claims for cost-effectiveness fail the axioms of representational measurement. Incremental cost-effectiveness ratios are only meaningful if both numerator and denominator satisfy the mathematical conditions required for ratio operations. Costs generally satisfy these requirements because monetary units possess ratio properties. The denominator, however, depends upon the validity of the QALY. If the QALY is not a ratio measure, the resulting ratio cannot possess lawful quantitative meaning regardless of the sophistication of subsequent modelling or statistical analysis. The interrogation therefore indicates that HAS accepts cost-effectiveness analysis as a methodological convention rather than as a conclusion justified by representational measurement.

The endorsement of QALY aggregation follows naturally from this position. Once QALYs are accepted as quantitative entities, aggregation across individuals, populations, interventions, and time horizons becomes routine. Yet aggregation cannot compensate for absent measurement. Repeated arithmetic operations do not strengthen scientific validity; they merely propagate the original measurement error throughout increasingly elaborate analyses. Lifetime simulations, budget impact models, probabilistic sensitivity analyses, and population projections therefore

become increasingly precise representations of quantities whose measurement properties remain unresolved.

The interrogation also reveals only limited endorsement of the proposition that non-falsifiable claims should be rejected while simultaneously accepting the proposition that reference-case simulations generate falsifiable claims. This finding has important implications for the philosophy of HTA. A simulation model projects assumptions regarding future events under specified structural conditions. Its outputs depend upon those assumptions. Such models are valuable tools for exploring scenarios and informing administrative decisions, but they do not constitute falsifiable scientific claims in the Popperian sense. A falsifiable claim requires prospective specification of the attribute, comparator, measurement framework, success criteria, and empirical conditions under which the claim may be refuted. Lifetime simulations cannot satisfy these requirements because they generate hypothetical projections rather than directly testable empirical propositions.

The interrogation further demonstrates weak recognition of the proposition that the outcome of interest for latent traits is possession of the latent attribute itself. This omission explains much of the confusion surrounding patient-reported outcomes within HTA. The objective is not to estimate questionnaire scores or preference weights but to estimate possession of the latent construct—pain, fatigue, functional limitation, treatment satisfaction, need fulfilment, or quality of life. Rasch measurement was developed precisely to estimate this possession while satisfying the requirements of invariant measurement. The absence of this concept from the HAS knowledge base means that latent outcomes continue to be represented through descriptive scoring systems rather than genuine measures.

Similarly, little recognition is given to the proposition that the Rasch rules for measurement are identical in purpose to the axioms of representational measurement. Both seek lawful quantitative representation of empirical reality. Representational measurement establishes the general conditions required for numerical assignment. Rasch theory provides the specific framework through which latent attributes satisfy those conditions. Their shared objective is measurement rather than statistical description. The interrogation indicates that this relationship receives almost no explicit attention within the HAS methodological framework, leaving latent measurement disconnected from the broader philosophy of quantitative science.

Taken together, these findings demonstrate a coherent pattern rather than isolated methodological preferences. The HAS knowledge base is exceptionally sophisticated in matters of evidence appraisal, comparative clinical assessment, health economics, reimbursement procedure, committee structure, uncertainty analysis, and policy evaluation. None of these achievements is questioned by the present interrogation. The issue lies elsewhere. The knowledge base consistently endorses numerical conventions while giving comparatively little recognition to the scientific principles that determine whether those numerical conventions constitute measurement. Measurement inversion is therefore not an incidental weakness but an organizing characteristic of the analytical framework itself.

The implications extend beyond HAS. As France occupies a central position within European health technology assessment, the methodological assumptions embedded within HAS guidance

inevitably influence university curricula, professional training, research centers, journal expectations, reimbursement dossiers, and methodological practice throughout the wider HTA community. Measurement inversion therefore becomes self-reinforcing. Successive generations of analysts inherit accepted procedures without exposure to the measurement principles necessary to evaluate them critically. The result is an intellectual environment in which methodological convention acquires the status of scientific truth simply through repetition.

The conclusion is therefore unavoidable. The interrogation demonstrates that the publicly accessible HAS knowledge base embodies a pervasive pattern of measurement inversion. Foundational propositions required by representational measurement receive consistently weak endorsement, while propositions supporting utilities, EQ-5D preference algorithms, QALYs, cost-effectiveness ratios, and reference-case simulation modelling receive consistently strong endorsement despite their incompatibility with the scientific conditions governing lawful quantitative claims. Arithmetic has displaced measurement as the organizing principle of HTA.

The significance of this finding extends beyond criticism of a single agency. HAS represents one of Europe's most influential HTA organizations. Its methodological guidance shapes professional expectations within France and contributes to the wider European HTA environment. If its analytical framework does not recognize the conditions required for quantitative measurement, then the consequences are necessarily systemic. The companion paper on curriculum inversion addresses the educational implications of this conclusion. If the scientific foundations of measurement are absent from the knowledge base, it becomes inevitable that future practitioners will be trained to reproduce the same methodological assumptions. Measurement inversion thereby becomes curriculum inversion, ensuring that the existing paradigm is transmitted from one generation of HTA professionals to the next. The challenge facing French HTA is therefore not incremental refinement of existing methods but reconstruction upon the foundations of representational measurement, linear ratio measurement for manifest attributes, Rasch logit ratio measurement for latent attributes, and evaluable, replicable, and falsifiable claims regarding therapy impact.

THE FUTURE OF HEALTH TECHNOLOGY ASSESSMENT IN FRANCE

The interrogation of the Haute Autorité de Santé (HAS) points to a clear and unavoidable conclusion: the French national HTA knowledge base is committed to measurement inversion. Statements that are false under the axioms of representational measurement are strongly endorsed, while statements that define the conditions for valid measurement are weakly endorsed or absent. This is not a marginal inconsistency or a local methodological weakness. It is a systematic framework in which numerical representation substitutes for measurement and arithmetic is applied to quantities whose measurement status has not been established.

The commitment to cost-utility analysis and QALY-based reasoning is explicit. Utilities are treated as if they possess interval or ratio properties, permitting multiplication by time, aggregation across individuals, incorporation into lifetime models, and use as denominators in cost-effectiveness ratios. These operations are foundational to contemporary HTA, yet they are not admissible unless the relevant quantities first satisfy the requirements of measurement. The HAS knowledge base gives little recognition to the requirement that multiplication demands ratio measurement, that

measurement must precede arithmetic, that latent attributes require Rasch-conformant measurement, or that unidimensionality, invariance, and lawful scale properties must be demonstrated before quantitative claims can be advanced. False measurement is not an exception within the framework. It is the rule.

What distinguishes the French case is the central role of HAS in establishing national methodological authority. HAS does not merely participate in HTA. It defines the framework through which manufacturers, consultants, academic researchers, committee members, and decision makers are expected to understand evidence, economic evaluation, comparative assessment, and reimbursement appraisal. If the HAS knowledge base embodies measurement inversion, then measurement inversion is not peripheral to French HTA. It is institutionalized at the national level.

This raises an unavoidable question regarding the future of HTA in France. Can a framework that is demonstrably inconsistent with the principles of measurement be sustained once this inconsistency is made explicit? For decades, authority has rested on the presence of numbers: utility values, QALYs, cost-effectiveness ratios, model outputs, probabilistic sensitivity analyses, and apparent methodological sophistication. The problem is that these numbers are not measures. They do not satisfy the conditions required for lawful quantitative claims. Once this is recognized, the scientific basis for their authority collapses.

The challenge is not technical but foundational. It is not a matter of refining models, improving real-world evidence, adjusting discount rates, expanding sensitivity analyses, or improving submission standards. These refinements leave the central problem untouched. A system that applies arithmetic to non-measures cannot generate evaluable claims. It cannot support replication in a measurement-consistent sense, nor can it be falsified as a scientific claim. The framework fails at the first step because the quantities to which arithmetic is applied have not been demonstrated to be measures.

The conclusion is therefore direct. French HTA cannot be reconstructed by improving the current reference-case apparatus. It must return to first principles. All therapy-impact claims must begin with the specification of the target attribute and must be grounded in valid measurement: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. These are the only defensible foundations for evaluable, replicable, and falsifiable claims regarding therapy impact. If HAS continues to rely upon utilities, QALYs, cost-effectiveness ratios, and reference-case simulations without first satisfying the requirements of representational measurement, the framework remains scientifically untenable. The future of HTA in France therefore depends not on methodological refinement but on reconstruction.

To facilitate this transition, Maimon Research LLC has developed a comprehensive nine-unit HTA Reconstruction Program ⁵. The program provides a systematic introduction to representational measurement, the theory of attributes, the principal scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, Rasch logit ratio measurement, protocol development, and the construction of evaluable, replicable, and falsifiable claims regarding therapy impact. Its purpose is not to modify the existing reference-case paradigm but to replace it with a scientific framework in which measurement once again precedes arithmetic.

The program has been designed for universities, HTA agencies, reimbursement organizations, research centers, professional societies, pharmaceutical companies, and health economists seeking a transition from assumption-driven modelling to scientifically defensible measurement. It provides a structured pathway for professional development while establishing the competencies required for the next generation of HTA practitioners. In this way, it offers not simply a critique of the existing paradigm but a practical route toward the reconstruction of HTA as a measurement-based scientific discipline.

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 of 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/>