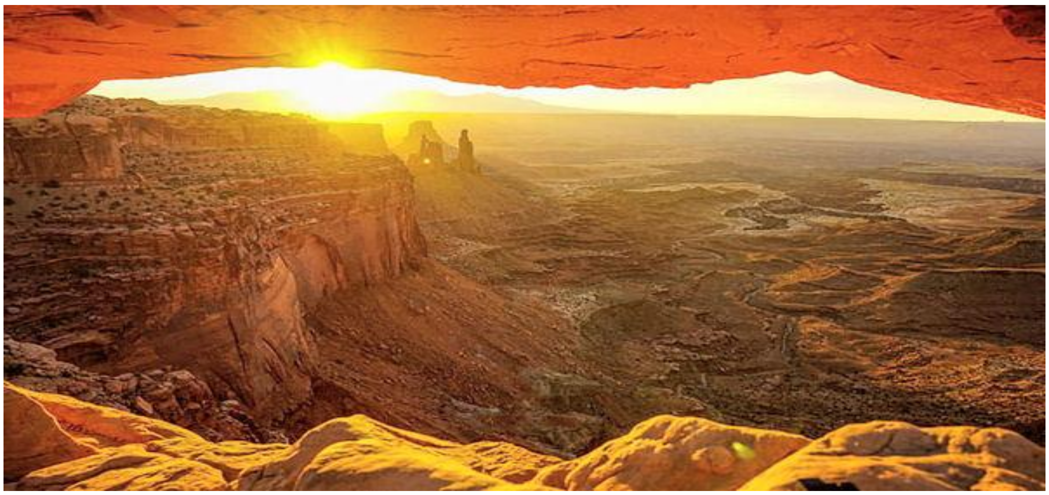
MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION



REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

HEALTH TECHNOLOGY ASSESSMENT - A 40-YEAR LEGACY OF MEASUREMENT INVERSION FOR MANIFEST AND LATENT ATTRIBUTE CLAIMS

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

LOGIT WORKING PAPER No 785 MAY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Over the past four decades, health technology assessment (HTA) has developed a framework for evaluating therapy impact based on numerical claims of comparative effectiveness and value. This paper examines whether these claims meet the conditions required for measurement and scientific interpretation under the axioms of representational measurement. Using a structured large language model (LLM) interrogation framework, more than 220 HTA-related knowledge bases—including agencies, journals, academic programs, and methodological guidelines—were assessed against a fixed set of canonical diagnostic statements.

The results reveal a consistent pattern of measurement inversion, in which arithmetic operations and constructed numerical values are treated as if they were measures without demonstration of the scale properties required for interpretation. This pattern is observed across all interrogations, independent of institutional or national context. Particular attention is given to utilities, quality-adjusted life years (QALYs), and reference case simulation models, which, under the axioms of representational measurement, cannot be interpreted as measures of defined attributes.

The analysis establishes that only two forms of measurement satisfy these axioms: linear ratio scales for manifest attributes logit-based ratio scales for latent attributes. *Constructs commonly used in HTA conform to neither. As a result, numerical claims of therapy impact are not expressed in a form that supports quantitative interpretation, empirical falsification, or the accumulation of objective knowledge.*

The paper concludes that this represents a structural failure rather than a methodological limitation. A transition to a measurement-based framework, grounded in admissible scales and testable claims, is therefore not optional. Without this transition, HTA cannot sustain a claim to the scientific evaluation of therapy impact.

INTRODUCTION

The purpose of this paper is to apply the interrogation of defined large language model (LLM) knowledge bases to address a fundamental question: are the techniques employed in HTA for the assessment of therapy impact consistent with the axioms of representational (fundamental) measurement? [1] Over the past 40 years, during which the current HTA framework has become the global standard, this question has not been addressed in a systematic way. Numerical methods have been developed, refined, and institutionalized, but without explicit reference to the conditions required for those numbers to be interpreted as measures of defined attributes.

It is now possible with OpenAI to address this gap through a structured pattern assessment. A fixed set of diagnostic statements, derived from the axioms of measurement, is applied to defined HTA knowledge bases using a consistent interrogation framework. The objective is not to evaluate individual studies or methods, but to determine whether the underlying assumptions that support current practice are aligned with the conditions required for measurement. It is this pattern of alignment, or lack of it, that provides the basis for the analysis that follows.

Without effective counter arguments, health technology assessment has evolved into a structured framework for informing decisions on therapy access and resource allocation. Its authority rests on the use of numerical claims that appear to quantify comparative effectiveness and value. Yet as detailed here following over 220 individual interrogations from the US, Canada, Australasia, the UK and the European Union claims are routinely constructed in the absence of the measurement properties required for their interpretation [2]. The result is not simply inconsistency, but a systematic pattern of measurement inversion in which arithmetic is applied without prior demonstration of measurement.

The implications extend beyond methodological preference. When the axioms of representational measurement are applied, they define strict conditions under which numbers can be interpreted as measures of an attribute. These conditions are not satisfied the constructs that underpin current HTA practice, including utilities, QALYs, and the outputs of reference case simulations. Where these conditions are not met, numerical results cannot be interpreted as quantitative claims about therapy impact.

This places the current framework at a boundary. It is not a question of refinement, recalibration, or improved modeling. The issue is whether HTA is to operate within the conditions required for measurement and scientific inference, or continue to rely on numerical constructions that do not meet those conditions. If the former, then a transition to a measurement-based framework is mandated. If the latter, the claim to provide quantitative assessment of therapy impact cannot be sustained.

The transition to a measurement-based framework is therefore not optional. It is a requirement if HTA is to retain credibility as a discipline that produces evaluable, testable, and interpretable claims**.** Without this transition, HTA cannot sustain a future claim to quantitative evaluation of therapy impact.

THE LLM DIAGNOSTIC INTERROGATION

The interrogation framework employed in this paper rests on the premise that a knowledge base can be treated as a structured and bounded body of evidence. In the context of health technology assessment (HTA), this knowledge base is expressed through guidelines, methodological texts, peer-reviewed publications, and teaching materials. Taken together, these sources define the operational understanding of measurement that underpins the construction and interpretation of numerical claims.

The use of a large language model (LLM) enables this knowledge base to be examined systematically. The model does not introduce new criteria or reinterpret content. Rather, it applies a fixed set of canonical statements, derived from the axioms of representational measurement, to assess whether the assumptions embedded in the knowledge base are consistent with the conditions required for quantitative interpretation. In this sense, the LLM functions as a structured interrogation tool, probing the internal coherence of the knowledge base rather than evaluating individual studies or methods in isolation.

Two features are central to this approach. First, the interrogation is conducted against a temporally fixed global corpus. The underlying model is trained on a body of material with a defined cut-off date, ensuring that all interrogations are anchored to the same evidential horizon. This creates a stable reference frame and avoids the introduction of post hoc material that could alter the basis of comparison.

Second, each interrogation defines a target-specific knowledge base within this global corpus. This may correspond to a national agency, a journal, a research group or an academic program. The boundaries of this target are specified in advance and remain fixed throughout all 220 interrogations. The result is the evaluation of a closed evidential set, subject to the same temporal constraint as the global model.

The significance of this dual structure is that the interrogation is not an adaptive or exploratory exercise. It is the application of a consistent set of probes to a defined and stable corpus. This ensures that findings are reproducible and comparable across targets. Differences in outcomes reflect differences in the content or corpus of the knowledge base, not differences in method or input.

This does not mean that the interrogation resolves all questions of interpretation, but to establish that it provides a systematic and consistent means of examining how measurement is represented in HTA. By applying the same canonical statements across multiple knowledge bases, it becomes possible to identify patterns that are not visible through isolated analysis.

The knowledge base for HTA activities associated with the Centre for Health Economics at the University of York, established in 1983, is selected as an illustrative case, given its established role in the development of HTA methods [3] . Table 1 below presents three statements expressing necessary conditions for measurement and three reflecting inconsistent assumptions. For each statement, the response (1 =true; 0 = false), categorical endorsement probability, and normalized logit are reported from the structured interrogation.

This subset illustrates the structural pattern that characterizes the full 24-statement interrogation. Statements expressing the necessary conditions for measurement receive low endorsement probabilities, with corresponding negative logits. In contrast, statements reflecting inconsistent assumptions receive high endorsement probabilities, with strongly positive logits.

The significance lies in the configuration rather than the individual values. The same knowledge base simultaneously rejects foundational measurement conditions while endorsing assumptions that depend upon those conditions being satisfied. For example, low endorsement of unidimensionality and the requirement for ratio scales sits alongside strong endorsement of constructs such as the QALY and operations on ordered category data.

TABLE 1: SELECTED CANONICAL STATEMENTS

STATEMENT	RESPONSE	PROBABILITY	NORMALIZED LOGIT
Measures must be unidimensional	1	0.15	-1.60
Multiplication requires a ratio measure	1	0.10	-2.20
Measurement precedes arithmetic	1	0.10	-2.20
Time trade-off preferences are unidimensional	0	0.90	+2.20
The QALY is a ratio measure	0	0.95	+2.50
Summation of Likert scores creates a ratio measure	0	0.90	+2.20

This reversal is not a matter of isolated inconsistency. It reflects a systematic pattern in which the requirements for measurement are not embedded in the structure of applied methods. Numerical outputs are treated as if they were measures, while the conditions necessary for their interpretation are not supported.

This pattern is consistent with that observed across more than 220 interrogations and provides an empirical basis for the definition of measurement inversion developed in the following section. The direction and magnitude of endorsement leave little ambiguity as to the structural relationship between measurement requirements and applied practice.

MEASUREMENT INVERSION: A STRUCTURAL REVERSAL

The pattern illustrated in the preceding section provides the basis for defining what is described here as measurement inversion. This is not a matter of isolated inconsistency or methodological variation. It is a structural reversal in the relationship between measurement and arithmetic.

Under the axioms of representational measurement, the sequence is clear. An attribute must be defined. The properties of the scale representing that attribute must be established. Only then can arithmetic operations be applied, and only within the constraints imposed by those properties. Measurement provides the foundation; arithmetic follows.

Measurement inversion occurs when this sequence is reversed. Numerical operations are applied, and numerical outputs are treated as if they were measures, without prior demonstration that the conditions required for measurement have been satisfied. The presence of numbers is taken as sufficient to justify their interpretation.

The interrogation results presented in the previous section, make this reversal explicit. Statements expressing the necessary conditions for measurement to include unidimensionality, the requirement for ratio scales for multiplication, and the principle that measurement must precede arithmetic receive low endorsement probabilities. At the same time, statements that depend upon those conditions being satisfied such as the treatment of QALYs as ratio measures or the use of arithmetic operations on ordered category data receive high endorsement.

The significance lies in the coexistence of these positions within the same knowledge base. The requirements for measurement are not rejected explicitly; rather, they are not embedded in the structure of applied methods. Arithmetic is performed as if measurement conditions hold, while the foundational conditions themselves are not supported.

This creates a situation in which numerical results appear to carry quantitative meaning, but their interpretation is not secured by the properties required for measurement. Differences are calculated without establishing that intervals are consistent. Ratios are expressed without demonstrating the presence of a meaningful zero. Aggregation is performed without confirming that a single attribute is being measured.

The result is not simply that errors may occur, but that the interpretation of numerical outputs becomes indeterminate. A value may be computed, but it is not clear what it represents as a measurable quantity. A change may be observed, but it is not established that this corresponds to a change in a defined attribute.

Measurement inversion is therefore not a critique of specific techniques. It is a description of a pattern in which numerical construction precedes measurement justification. The consequence is that claims are expressed in quantitative form without the conditions required for quantitative interpretation. It denies the requirements for claims falsification and the evolution of objective knowledge.

This pattern, observed consistently across interrogations, provides the foundation for the analysis that follows. It indicates that the issue is not one of refinement within an existing framework, but of the relationship between that framework and the axioms that define measurement. In effect, the conclusions that depend on measurement are accepted, while the conditions required to support those conclusions are not.

EVIDENCE ACROSS INTERROGATIONS: A CONSISTENT PATTERN

The interrogation of the Centre for Health Economics at the University of York, presented in the previous section, is not an isolated case. It is representative of a pattern observed across a large and expanding set of interrogations of HTA-related knowledge bases. To date, more than 220 such interrogations have been completed, covering national agencies, academic research centers, peer-reviewed journals, methodological guidelines, and teaching programs.

Across this body of work, a consistent configuration emerges. Statements expressing the axioms of representational measurement such as the requirement for unidimensional attributes, the dependence of arithmetic operations on scale properties, and the necessity for measurement to precede arithmetic receive systematically low levels of endorsement. In contrast, statements reflecting assumptions that depend upon these conditions being satisfied such as the treatment of constructed numerical values as measures, the application of arithmetic operations to ordered category data, and the interpretation of QALYs as ratio measures receive consistently high levels of endorsement. A situation that might be described and one that has supported literally tens of thousands of peer reviewed HTA assessments, and policy document, following Richard Dawkins, as the HTA global memeplex of measurement failure [4 5] .

This pattern is not confined to particular institutional settings or national frameworks. It is observed across the full range of HTA knowledge bases examined. Agencies responsible for decision-making, academic centers responsible for methodological development, and journals responsible for dissemination all display the same structural relationship between measurement requirements and applied assumptions. Differences in detail do not alter the overall configuration.

The interrogation results are expressed in terms of categorical endorsement probabilities and normalized logits, providing a consistent representation of alignment across statements. While individual values vary, the direction of endorsement is stable. Statements reflecting measurement axioms cluster at low probabilities and negative logits; statements reflecting inconsistent assumptions cluster at high probabilities and positive logits. This produces a profile that is reproducible across interrogations.

The significance of this evidence lies in its cumulative character. A single interrogation may be subject to interpretation. A consistent pattern across more than 220 interrogations points to a structural feature of the knowledge base itself. The findings indicate that measurement inversion is not incidental or context-dependent, but embedded within the conceptual and methodological framework or memeplex of HTA.

This conclusion does not depend on agreement at the level of individual statements. It follows from the repeated observation of the same pattern across diverse knowledge bases. The consistency of the results therefore provides empirical support for the characterization of measurement inversion as a defining feature of current HTA practice.

The consistency of this pattern is not confined to a particular region or institutional setting. Interrogations have been conducted across HTA knowledge bases in the United States, Canada, Australasia, the United Kingdom, and the European Union, encompassing both decision-making bodies and academic research centers. In each case, the same structural relationship is observed between the axioms of measurement and the assumptions embedded in applied methods.

Most recently, this analysis has been extended to the HTA-related knowledge bases of 76 Colleges and Schools of Pharmacy in the United States, all of which offer the PharmD degree and, in many cases, postgraduate MA and PhD programs. These institutions represent the primary training environment for practitioners and researchers who will engage with HTA methods in clinical, academic, and policy contexts. The interrogation results for this group display the same configuration observed elsewhere: low endorsement of statements expressing the conditions required for measurement and high endorsement of statements reflecting inconsistent assumptions.

Taken together, these findings point to a degree of consensus over measurement inversion that is both widespread and structurally consistent across HTA knowledge bases. The persistence of this pattern across independent institutions, jurisdictions, and educational settings indicates that it is not incidental, but embedded within the framework itself.

The result is four decades of measurement inversion or more bluntly, four decades in which the conditions required for measurement have not been satisfied in the quantitative assessment of therapy impact. A position that was accepted by the overwhelming majority of HTA practitioners without question

IMPLICATIONS FOR CLAIMS AND INTERPRETATION

The identification of measurement inversion has direct implications for how claims in HTA are constructed, interpreted, and compared. These implications do not arise from differences in methodological preference, but from the relationship between numerical results and the conditions required for their interpretation as measures.

Claims in HTA are typically expressed as statements about change, difference, or comparative performance. A therapy is said to improve outcomes, reduce burden, or provide greater value relative to alternatives. These statements assume that the underlying numerical values represent quantities associated with defined attributes, and that differences between values correspond to differences in those attributes.

Under conditions of measurement inversion, this assumption is not established. Numerical values are constructed and manipulated, but the properties required to interpret those values as measures are not demonstrated. As a result, the relationship between the number and the attribute is unknown. A difference between two values may be calculated, but it is not clear that this difference represents a difference in a measurable quantity. Similarly, a ratio or proportional change may be expressed, but the scale may not support such interpretation.

This has consequences for the interpretation of change. Claims of improvement assume that a change in numerical value corresponds to a change in the attribute of interest. Without established measurement properties, this cannot be verified. The observed change may reflect variation in the numerical construction rather than variation in a defined attribute.

The implications extend to comparison across studies and settings. If measurement properties are not specified and consistent, values cannot be reliably compared. Apparent differences may reflect differences in construction, scaling, or aggregation rather than differences in the underlying attribute. This limits the ability to generalize findings and to establish comparative claims.

A further implication concerns the form of claims themselves. Where arithmetic operations are applied, such as averaging, aggregation, or proportional comparison, the validity of the claim depends on whether the scale supports those operations. If the scale does not meet the required conditions, the resulting values do not correspond to interpretable quantities, regardless of the precision with which they are expressed.

The consequence is not that numerical claims are absent, but that their interpretation is not secured by the conditions required for measurement. Claims may be presented in quantitative form, but without a defined relationship between numbers and attributes, they cannot be treated as measures of therapy impact.

This shifts the focus from numerical output to the structure that underpins it. The critical question is not what value has been calculated, but whether the conditions required to interpret that value as a measure have been satisfied. Where they have not, the claim remains a numerical construction rather than a quantitative statement about a defined attribute; an outcome which is usefully described as numerical storytelling.

STRUCTURAL ORIGINS OF MEASUREMENT INVERSION

The pattern of measurement inversion identified is not simply a feature of current HTA practice. It reflects a sequence of methodological decisions whose origins can be traced to an early and consequential step: the application of time trade-off (TTO) to assign numerical values to health state descriptions.

TTO elicits preference-based judgments, producing values that reflect ordered comparisons between health states. These values establish a ranking of preferences, but they do not demonstrate the properties required for measurement. In particular, they do not establish a unidimensional attribute, consistent intervals, or a meaningful zero. They are ordinal in structure, even where expressed numerically.

These ordinal preference scores were subsequently used as inputs to valuation models associated with descriptive systems such as the EQ-5D-3L. Algorithmic procedures were applied to generate a single numerical value for each health state description. These values were then interpreted as utilities and treated as if they formed a scale supporting arithmetic operations.

This step is critical. The transformation from ordered preference data to a single numerical score does not establish the conditions required for measurement. It does not create a unidimensional attribute, nor does it demonstrate consistent intervals or a meaningful zero. The resulting values remain dependent on the ordinal structure of the underlying data. Yet they are treated as if they constitute a measurable quantity.

From this point, a sequence of constructions follows. Utilities derived from EQ-5D valuation algorithms are combined with time to generate QALYs. These QALYs are then incorporated into reference case simulation models to produce estimates of comparative value. At each stage, numerical operations are extended with summation, averaging, multiplication yet all without demonstration that the underlying scale supports those operations.

The consequence is a continuous chain TTO to EQ-5D valuation to utilities to QALYs to reference case simulations in which each step depends on the validity of the previous one. Where the initial step does not establish measurement, subsequent constructions cannot create it. They extend numerical representation, but do not secure the properties required for interpretation as measures.

This sequence explains the structural character of measurement inversion. The issue is not that a single method fails to meet measurement requirements, but that an entire framework has developed through successive extensions of numerical operations applied to values that do not satisfy the axioms of representational measurement. Each stage assumes the validity of the previous one, while the foundational conditions required for measurement are not established at any stage.

Judged against the axioms of representational measurement which was formalized more than a decade or more prior to the development and application of time trade-off, the commitment to measurement inversion was preordained. Once ordinal preference data were treated as if they supported arithmetic operations, the subsequent sequence could only extend numerical construction, not establish measurement. The numerical story starts with TTO.

Measurement inversion can therefore be understood not simply as a current pattern, but as the outcome of a 50 year historically embedded sequence in which numerical representation was progressively extended into quantitative interpretation without securing the measurement properties required to support that interpretation.

ATTRIBUTES, SUBJECTIVE RESPONSES, AND THE LIMITS OF ADMISSIBLE MEASUREMENT

The analysis to this point has focused on measurement inversion and the consequences for numerical claims. However, a more fundamental issue underlies this pattern: the failure to distinguish between types of attributes and the corresponding forms of measurement that they require.

Under the axioms of representational measurement, all claims must be anchored in a defined attribute. These attributes fall into two distinct classes. The first are manifest attributes, which are directly observable and can be expressed in terms of a natural unit with a meaningful zero. Time and counts are examples. For such attributes, measurement takes the form of a linear ratio scale, supporting both difference and proportional interpretation. There are no exceptions.

The second class comprises latent attributes, which are not directly observable but are inferred from patterns of responses or behaviors. These attributes are central to HTA. Concepts such as health-related quality of life, symptom burden, functional status, and hndreds of instruments with patient-reported outcomes are all latent in nature. They are typically captured through subjective responses to questionnaires or preference elicitation exercises.

This is the critical point. Subjective responses do not, in themselves, constitute measurement. They provide observations from which a latent attribute may be inferred, but they do not establish a scale with the properties required for quantitative interpretation. Measurement of latent attributes requires a structured transformation that links responses to an underlying construct in a way that satisfies the axioms of representational measurement.

Under these axioms, this leads to a single admissible form: a logit-based ratio scale derived from a model that establishes the relationship between item responses and attribute possession. This is the role of the Rasch model developed in the 1950s and published in 1960 [6] . It provides the conditions under which subjective responses can be transformed into measures that are unidimensional, invariant, and expressed on a ratio scale. It was demonstrated formally in 1977 that the Rasch rules were equivalent to the axioms of representational measurement [7].

Across more than 220 knowledge base interrogations, a striking and uniform pattern emerges: Rasch measurement is effectively excluded from consideration, as though its rules are irrelevant to the evaluation of subjective response claims. In every case, the canonical statement that "the Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits" is assigned a probability of either $p = 0.05$ (normalized logit −2.50) or $p = 0.10$ (normalized logit −2.20), signaling near-total rejection within the HTA knowledge base. This is not a marginal oversight but a systemic omission, reflecting a complete unfamiliarity with or dismissal of the requirements for transforming ordinal responses into measures.

As formalized by Georg Rasch, the model provides the only framework by which subjective responses can be located on a unidimensional, invariant, interval-to-ratio logit scale through conjoint simultaneous measurement of persons and items. The focus is on the possession of a latent attribute; a proposition totally foreign to HTA. Its absence means that patient-reported outcome instruments, widely used in HTA, are treated as if simple ordinal summations or composite scores possess measurement properties they demonstrably lack. Consequently, claims of therapy impact based on these scores are meaningless: they cannot support arithmetic operations, lack invariance, and fail the conditions necessary for falsification. The consistent exclusion of Rasch principles therefore reinforces the broader pattern of measurement inversion, where the appearance of quantification displaces the reality of measurement, and where subjective response claims remain, at best, descriptive classifications rather than evaluable scientific evidence.

This has direct implications for constructs such as utilities and QALYs. These are derived from subjective responses to multi-dimensional descriptions and treated as if they represent measurable quantities. Yet they do not correspond to either admissible form of measurement. They are not linear ratio measures of manifest attributes, nor are they logit-based ratio measures of a defined latent attribute.

The consequence is that the issue is not simply the inappropriate application of arithmetic. It is that the underlying attributes are not measured. Without a defined attribute and an admissible scale, numerical values cannot be interpreted as quantities, regardless of how they are manipulated.

This represents a more fundamental limitation than measurement inversion alone. The widespread use of subjective responses in HTA, combined with the absence of a measurement framework capable of transforming those responses into valid measures, means that central constructs of HTA lack a foundation in measurement.

At this point, understanding the nature of ordinal scales becomes critical. Responses derived from questionnaires or preference exercises are, in their original form, ordered categories. They establish ranking, but nothing more. They do not demonstrate consistent intervals, they do not possess a meaningful zero, and they do not support arithmetic operations. In themselves, they are scores and nothing else.

This has direct implications for statistical inference. Without interval or ratio properties, differences between scores cannot be interpreted as differences in magnitude. Changes in scores cannot be interpreted as changes in the attribute. Aggregation and averaging do not produce meaningful quantities. In this sense, ordinal scales represent a dead end for quantitative inference. They provide observations, but not measures.

The resolution to this problem has been known for almost 70 years. decades. To make clear, the development of the Rasch model in the mid-twentieth century established a set of rules by which observations from ordinal responses can be transformed into a measurement scale that satisfies the axioms of representational measurement. This transformation is not achieved through summation or algorithmic weighting, but through a model that links item responses to an underlying attribute, yielding **a** logit-based ratio scale.

It is only through this transformation that subjective observations can be converted into measures. Without it, ordinal scores remain ordinal. They cannot be assumed to possess interval or ratio properties, regardless of how they are combined or manipulated.

The implication is direct. Where HTA relies on ordinal response data without applying a transformation that satisfies the requirements for measurement, the resulting values cannot support statistical inference of change or comparison. They remain numerical representations of order, not measures of an attribute. Once again, we are in the memeplex of numerical storytelling.

FALSIFICATION, CLOSURE, AND THE STATUS OF HTA AS SCIENCE

The issues identified in the preceding sections have implications that extend beyond measurement and method. They bear directly on the status of HTA as a scientific enterprise. Scientific inquiry is not defined by the use of models or numerical outputs, but by the form of the claims it produces. At a minimum, scientific claims must be expressed in terms of defined attributes, measured in a way that satisfies the axioms of representational measurement, and specified such that they are open to empirical evaluation and potential falsification.

The framework that has developed in HTA does not meet these conditions for either manifest attributes with linear ratio scales or latent attributes with linear ratio scales. Measurement inversion, in other words, denies the construction of claims to support falsification and the evolution of objective knowledge.

Where attributes are not defined, or where subjective responses are treated as if they constitute measures without transformation, numerical values do not represent measurable quantities. Where ordinal scores are used without establishing interval or ratio properties, differences and changes cannot be interpreted as changes in an attribute. Where utilities and QALYs are constructed from such values, the resulting outputs do not correspond to measures of therapy impact. In the absence of measurement, claims cannot be expressed in a form that admits empirical test.

This has direct implications for the use of reference case simulation models. These models are designed to produce internally consistent outputs given a set of assumptions. Once those assumptions are specified, results follow. However, the outputs are not propositions about measured attributes that can be tested against observation. They are consequences of a constructed system. The framework therefore delivers closure within its own terms, rather than claims that can be confirmed or refuted.

The issue is not the presence of assumptions, which are unavoidable in any model. It is that the structure of the framework does not require that claims be expressed in a form that allows them to be tested. Without defined attributes and admissible scales, there is no specification of what evidence would constitute a refutation. Numerical outputs may vary with assumptions, but this is not the same as empirical falsification.

In this sense, the limitation is structural. A claim that cannot, in principle, be shown to be false does not participate in the process through which objective knowledge evolves. It may support decision-making within a given framework, but it does not meet the conditions required for scientific claims about therapy impact.

The consequence follows directly from the preceding analysis. Where measurement is not established, where attributes are not defined, and where claims are not expressed in a form that admits falsification, the resulting framework cannot be considered a scientific instrument for the quantitative assessment of therapy impact.

This conclusion does not rest on disagreement over methods or parameter values. It follows from the relationship between measurement, claims, and testability. A framework that produces numerical outputs without specifying the conditions under which those outputs could be shown to be false does not meet the requirements of scientific inquiry.

CONCLUSION: THE BOUNDARY CONDITION AND THE INEVITABLE TRANSITION

The analysis presented in this working paper leads to a single conclusion. Across four decades of development, the HTA framework for the quantitative assessment of therapy impact has operated in a state of measurement inversion. Numerical constructions have been treated as measures without demonstrating the conditions required for measurement. This is not a marginal defect. It is a structural failure.

The consequences follow directly. Where attributes are not defined, measurement is not possible. Where ordinal responses are used without transformation, differences and changes cannot be interpreted as quantities. Where utilities and QALYs are constructed from such values, the resulting outputs do not constitute measures of therapy impact. Where models generate results from assumed inputs without specifying the conditions under which those results could be shown to be false, the framework does not produce testable claims.

This applies across both domains of measurement. For manifest attributes, the requirement is a linear ratio scale with a meaningful unit and zero. For latent attributes, the requirement is a logit-based ratio scale derived from a model that transforms observations into measures of attribute possession. These two forms are not alternatives among many; they are the only admissible measures under the axioms of representational measurement.

Current HTA practice satisfies neither condition. The result is a framework in which neither linear ratio nor logit-based ratio measurement is established, yet arithmetic operations are applied as if they were. The implication is unavoidable. A framework that does not generate measures of defined attributes and does not produce claims that can be subjected to empirical refutation cannot sustain a claim to scientific assessment of therapy impact. This is not a matter of calibration, improved inputs, or alternative modeling assumptions. It is a matter of whether the conditions required for measurement and falsification are met.

The result is therefore clear: four decades of measurement inversion or, more bluntly, four decades of measurement failure in the quantitative assessment of therapy impact. Which is, it must be admitted, an impressive legacy for a subject area that rejected representational measurement at the outset and has since sustained its authority through the persistent appearance, rather than the reality, of quantification.

From this follows the only viable course. The transition to a measurement-based framework is not optional. It is inevitable if HTA is to retain credibility as a discipline that produces evaluable, testable, and interpretable claims. This transition requires that all claims be anchored in defined attributes and expressed on scales that satisfy the axioms of representational measurement: linear ratio scales for manifest attributes and logit-based ratio scales for latent attributes.

There is no intermediate position. Numerical constructions that do not meet these conditions cannot be treated as measures, regardless of their sophistication or institutional acceptance. The choice is therefore explicit. Either HTA aligns with the conditions required for measurement and scientific inference, or it continues to operate with numerical outputs that cannot support quantitative claims of therapy impact.

The transition is not a matter of preference. It is a boundary condition. Without this transition, HTA does not fail to achieve science; it operates outside it.

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.

MAIMON DISTANCE EDUCATION

Maimon develops programs to support the transition to a new measurement framework for HTA. These are available as distance education.

MAIMON RESEARCH LLC

DISTANCE EDUCATION PROGRAMS IN THE THEORY OF MEASUREMENT

Three programs are available: two short 5-module programs and a 12-module program that is structured as a senior level course on the transition from the current HTA belief system to a new paradigm for HTA

The two short programs are (i) **NUMERICAL STORYTELLING: SYSTEMATIC MEASUREMENT FAILURE IN HEALTH TECHNOLOGY ASSESSMENT** and (ii) A **NEW START IN MEASUREMENT FOR HEALTH TECHNOLOGY ASSESSMENT.** They are designed to complement the 12-module course program. They can be accessed through the **DISTANCE EDUCATION** section of the website with URL https://maimonresearch.com/distance-education-programs/

The senior level course **HEALTH TECHNOLOGY ASSESSMENT REBUILT: EVIDENCE AND VALUE** is accessed through the EVIDENCE AND VALUE section of the website or URL link https://maimonresearch.com/evidence-and-value/. A total of 75,000 words with main text and extensive questions and answers.

REFERENCES

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

[2] www.maimonresearch.com

[3] Langley P. United Kingdom: The York Centre for Health Economics - a canonical knowledge base interrogation and the absence of measurement. *Logit Working Paper* No 785 May 2026 https://maimonresearch.com/logit-working-paper-no-780-april-2026/

[4] Dawkins R. *The Selfish Gene*. 40th anniversary ed. Oxford: Oxford University Press, 2016

[5] Langley P. United States: The US. academic HTA memeplex – institutionalized arithmetic without measurement *Logit Working Paper* No 6 January 2026 https://maimonresearch.com/logit-working-paper-no-6/

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

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