MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

THE MEASUREMENT FUTURE IN HEALTH TECHNOLGY ASSESSMENT

LEGACY HTA: ISPOR – "VALUE IN HEALTH" ENDORSES MEASUREMENT INVERSION

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

LOGIT WORKING PAPER No 2001 AUGUST 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Value in Health, *the flagship journal of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), occupies a central position in the development and dissemination of health economics, outcomes research and health technology assessment (HTA). This paper interrogates its cumulative knowledge base against 24 canonical propositions derived from representational measurement theory and Rasch measurement principles. Categorical endorsement probabilities, transformed to normalized logits, assess whether the journal knowledge environment possesses the measurement requirements necessary to justify the arithmetic of contemporary HTA.*

The results identify a pronounced pattern of measurement inversion. Foundational requirements are weakly possessed: measurement preceding arithmetic receives an endorsement probability of 0.10 (−2.20 logits), multiplication requiring ratio measurement 0.10 (−2.20), and the requirement that representational measurement axioms precede arithmetic 0.10 (−2.20). By contrast, the knowledge base strongly reinforces practices inconsistent with these requirements. The false propositions that the QALY is a ratio measure and that QALYs can be aggregated each receive 0.95 (+2.50), while the false proposition that ratio measures can have negative values receives 0.90 (+2.20). Rasch measurement requirements for latent attributes occupy the negative extreme of the profile.

The findings are not unique to Value in Health. *Interrogations of other major journals in health economics, pharmacoeconomics, outcomes research and HTA demonstrate essentially the same structure. What has been replicated across the journal literature is the methodology, not the measurement proof. Peer review, repeated publication, professional consensus and increasing methodological sophistication cannot establish measurement properties absent before arithmetic begins.*

The implications for Value in Health *are therefore immediate. As ISPOR's principal journal, it has been instrumental in publishing and legitimizing utilities, QALYs, ICERs, cost-effectiveness analysis and reference-case modelling. If pervasive measurement inversion requires abandonment of legacy HTA, its leading journal cannot remain scientifically unchanged. There is no middle ground in which inadmissible arithmetic becomes admissible through familiarity, usefulness or professional endorsement. The question now confronting* Value in Health *is whether it continues refining the inherited framework or requires quantitative therapy assessment to begin with lawful measurement. Measurement must precede arithmetic.*

INTRODUCTION

Value in Health occupies a distinctive position within the international development of health economics, outcomes research and health technology assessment (HTA). As the flagship journal associated with the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), it has provided a major publication environment for economic evaluation, cost-effectiveness and cost-utility analysis, decision modelling, health-state preference measurement, patient-reported outcomes, value assessment and related methodological research. Its significance extends beyond the individual studies appearing in its pages. Through publication, methodological debate and the repeated application of accepted analytical conventions, *Value in Health* contributes to defining the intellectual environment within which contemporary HTA is conducted.

This makes the journal an important complement to the assessment of ISPOR Good Practices. Good Practice reports formally recommend, codify and transmit methodological standards. A scientific journal performs a different but equally important function. It provides the environment in which those standards are applied, developed, scrutinized and reproduced. Methods repeatedly appearing in a leading journal acquire professional legitimacy through use. They become familiar to researchers, enter academic teaching, inform manufacturer analyses and consultancy, provide precedents for subsequent publications and contribute to the methodological expectations of HTA agencies. A journal is therefore not simply a repository of research findings. It is part of the intellectual infrastructure through which a discipline defines acceptable scientific practice.

The fundamental question is: does the journal's cumulative knowledge base recognize and apply the measurement requirements necessary to determine whether the arithmetic underlying legacy HTA claims for therapy impact is admissible?

In the physical sciences and the more advanced social science disciplines, quantitative science begins with measurement rather than arithmetic [1 2]. Before numerical values are added, subtracted, multiplied, divided or incorporated into statistical and decision models, it must first be established that they represent defined attributes and possess the scale properties required for the proposed operations. Nominal scales classify, ordinal scales establish order, interval scales provide equal intervals without a non-arbitrary zero, and ratio scales possess both equal units and a non-arbitrary zero. Multiplication and division require ratio measurement. These restrictions are not optional methodological preferences. They determine whether arithmetic performed upon numerical representations has empirical meaning.

These requirements have immediate implications for the literature represented in *Value in Health*. Cost-utility analysis constructs the quality-adjusted life year (QALY) through multiplication, QALY = U × T. Time can satisfy the requirements for linear ratio measurement. The measurement status of the utility value is therefore decisive. If utility is to operate as a dimensionless proportional adjustment to time, the required measurement properties must be established independently before multiplication occurs. They cannot be created by the multiplication itself.

The problem becomes particularly evident when health-state utility values can take negative values while simultaneously being treated as though they support ratio operations. A ratio measure requires a non-arbitrary zero representing absence of the measured attribute. The presence of values below that zero is incompatible with the proportional interpretation required for ratio measurement. Naming a numerical health-state value a utility, anchoring it at selected points or generating it through a preference algorithm cannot establish the required measurement structure.

The consequences extend throughout the subsequent arithmetic. If utility has not been demonstrated to support multiplication by time, calculating a numerical product does not create a legitimate QALY measure. Aggregating QALYs cannot repair the initial failure. Subtracting them to generate incremental QALYs cannot establish missing measurement properties, and dividing incremental costs by an inadmissible ΔQALY cannot produce a scientifically meaningful cost-per-

QALY ratio. Increasing the sophistication of the subsequent analysis cannot change this conclusion. Probabilistic sensitivity analysis, microsimulation, scenario analysis, model validation and increasingly elaborate statistical procedures cannot manufacture measurement properties that were absent before the arithmetic began.

The same distinction applies to patient-reported outcomes, another major component of the research environment represented in *Value in Health*. Pain, fatigue, physical function, treatment satisfaction, need fulfilment and related concepts are latent attributes. Assigning integers to ordered response categories and summing them generates a score; it does not by itself construct a measure. Reliability, validity, responsiveness and statistical association do not establish equal units, invariance or ratio properties. A measurement system for a latent attribute must first establish a defined unidimensional construct and an invariant quantitative continuum upon which possession of that attribute can be located. Within the framework applied in this assessment, Rasch measurement provides the required person-item conjoint measurement pathway [3, 4].

The distinction between numbers, scores and measures is therefore central to the present interrogation. A journal may contain extensive numerical analysis while providing little recognition of the measurement conditions governing that analysis. Indeed, analytical sophistication can obscure the distinction. Once a numerical representation has become familiar and has been repeatedly subjected to statistical analysis, modelling and publication, its status as a legitimate quantitative measure can appear self-evident. But repeated arithmetic cannot retrospectively establish measurement.

This possibility is particularly important in the case of *Value in Health* because journal publication introduces a mechanism of methodological reproduction different from formal practice guidance. Individual papers are produced by different authors, subjected to editorial and peer-review processes and subsequently incorporated into the published scientific literature. No inference should therefore be made that individual authors, reviewers or editors consciously endorse the measurement assumptions examined in this assessment. The unit of analysis is the cumulative publicly represented knowledge environment of the journal, not the beliefs or competence of individuals.

The scientific question concerns what has survived that collective publication process. If utilities, QALYs, summated scores, cost-effectiveness ratios and reference-case simulations recur extensively while the measurement requirements necessary to justify those operations remain weakly represented, then measurement inversion has survived not merely methodological convention but the normal processes of scientific publication. Peer review may assess model structure, statistical methods, parameter selection, uncertainty, reporting and interpretation while leaving logically prior questions of measurement unresolved.

This distinction also bears upon falsifiability. Decision models generate outputs conditional upon assumptions, model structures and parameter values. Sensitivity and scenario analyses can examine the consequences of varying those assumptions, but internal variation of a model is not equivalent to empirical falsification. A scientific literature can therefore be apparently highly sophisticated in modelling uncertainty while remaining weakly connected to the requirement that therapy-impact claims should ultimately be empirically evaluable, replicable and capable of falsification.

The historical importance of *Value in Health* makes this interrogation especially consequential. The journal has contributed substantially to methodological development while disseminating many of the conventions that now define legacy HTA. Earlier examination of its knowledge environment suggested precisely this contrast: considerable attention to methods accompanied by comparatively limited attention to the measurement foundations required to determine whether the quantities entering those methods are legitimate measures [5].

The present assessment examines that problem systematically using the same 24 canonical propositions applied to the ISPOR Good Practices knowledge base and the wider program of HTA knowledge-base interrogations. Each proposition has a defined TRUE or FALSE status under representational measurement theory and the measurement framework employed here. Categorical endorsement probabilities assess the extent to which the *Value in Health* knowledge environment explicitly or implicitly reinforces each proposition, with those probabilities transformed to normalized logits to reveal the relative possession or non-possession of the underlying measurement principles.

A high endorsement probability for a FALSE proposition does not imply that authors, reviewers or editors knowingly believe a false statement. It indicates that the cumulative literature behaves as though that proposition were acceptable through the methods, assumptions and arithmetic repeatedly represented within it. Conversely, low endorsement of a TRUE proposition indicates that the measurement principle does not function strongly within the journal's publicly represented knowledge environment.

The question addressed is therefore narrower than an assessment of the overall scientific contribution of *Value in Health*, but more fundamental to its quantitative foundations. Does the knowledge base of a leading international journal in health economics and outcomes research begin where quantitative science must begin—with measurement—or has it helped normalize a literature in which numerical representation is accepted as measurement and arithmetic is permitted to precede the demonstration of the properties required to make that arithmetic admissible?

If the latter is demonstrated, measurement inversion has become embedded not merely in HTA guidelines or agency reference cases, but in one of the principal publication environments through which the discipline creates, evaluates and reproduces its scientific knowledge.

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 *VALUE IN HEALTH* KNOWLEDGE BASE

For the purposes of this assessment, the *Value in Health* knowledge base is defined as the cumulative methodological environment represented by the research, methodological papers, reviews, commentaries and related scientific discourse published in the journal. It is not identified with any single article, author, editorial decision or period of publication. Rather, it represents the recurring concepts, assumptions, analytical practices and quantitative conventions through which the journal has contributed to the development and communication of health economics, outcomes research and health technology assessment (HTA).

As the flagship journal associated with ISPOR, *Value in Health* occupies an important position within the international HTA knowledge environment. Its published literature encompasses economic evaluation, cost-effectiveness and cost-utility analysis, decision-analytic modelling, preference-based health-state assessment, patient-reported outcomes, value assessment, uncertainty analysis and related methodological research. These domains collectively represent much of the quantitative architecture of contemporary HTA.

A central feature of this knowledge base is its extensive use of numerical representations. Costs, probabilities, survival estimates, durations, health-state values and other inputs are combined within economic and decision models. Preference-based health-state values are treated quantitatively and combined with time to generate quality-adjusted life years (QALYs). QALYs are accumulated across health states and time periods, differences between treatment alternatives calculated and incremental outcomes related to incremental costs. These operations have become part of the established analytical language represented throughout the journal.

Preference-based multiattribute instruments constitute an important component of this environment. Responses across several health dimensions define health states; preferences for those states are elicited or estimated; and algorithms assign numerical values that can subsequently be incorporated into economic evaluations as utilities. Once admitted to cost-utility models, these values are treated as quantities capable of multiplication by time, aggregation across periods and comparison between treatment alternatives.

Patient-reported outcomes constitute another substantial component of the knowledge base. Responses to ordered categories may be assigned numerical values, combined into scores, transformed statistically and used to characterize health status, symptoms, functioning, quality of life and treatment effects. Instrument development consequently emphasizes concepts such as reliability, validity, responsiveness, interpretability and statistical performance. The resulting numerical scores can then enter statistical analyses, clinical comparisons and value assessments.

Decision-analytic modelling provides a further major component. Models combine evidence from multiple sources, extrapolate beyond observed study periods and estimate outcomes under alternative treatment strategies. Deterministic sensitivity analysis, probabilistic sensitivity analysis and scenario analysis are employed to explore uncertainty. Model structure, parameterization, validation, transparency and reporting have consequently become important subjects within the journal's methodological literature.

Taken together, these areas constitute an extensive and technically sophisticated quantitative knowledge environment. The journal has devoted considerable attention to the development, refinement and application of analytical methods. The issue addressed in the present interrogation is not whether *Value in Health* possesses such methods. It clearly does. The question is whether the knowledge base also possesses the logically prior measurement principles required to determine whether the numerical representations entering those methods are measures and whether the arithmetic performed upon them is admissible.

This distinction provides the boundary for the interrogation. A preference algorithm producing a numerical value does not by itself establish an interval or ratio measure. Assigning integers to ordered patient responses and summing them creates a score but does not demonstrate equal units. Anchoring a health-state value at zero does not establish a non-arbitrary zero. Multiplying a health-state value by time does not demonstrate that the value possessed the ratio properties required for multiplication before the operation occurred. Statistical transformation, validation and repeated publication cannot independently confer measurement properties.

The knowledge base is therefore interrogated for the extent to which it distinguishes between numbers, scores and measures. This is particularly important because a mature scientific literature can become increasingly sophisticated in the manipulation of numerical representations while leaving their measurement status implicit. Once a numerical convention is routinely employed, published and cited, its quantitative legitimacy can become assumed rather than demonstrated.

For manifest attributes such as survival time, hospital days, treatment duration or walking distance, the relevant requirement is an appropriate linear ratio measurement structure with equal units and a non-arbitrary zero. For latent attributes such as pain, fatigue, physical function or need fulfilment, responses must support the construction of a defined unidimensional attribute and an invariant quantitative continuum upon which possession of that attribute can be located. Within the measurement framework applied in this assessment, Rasch measurement provides the relevant pathway for latent attributes.

The distinction extends to model-based claims. Sensitivity analysis can determine how model outputs respond to changes in assumptions or parameter values, but it does not establish the measurement properties of the inputs. Similarly, internal and external model validation address aspects of model performance without converting conditional simulations into independently falsifiable empirical claims. Measurement and falsifiability therefore remain logically prior scientific requirements rather than consequences of modelling sophistication.

The unit of analysis is explicitly the cumulative journal knowledge environment rather than individual authors, reviewers or editors. Individual papers address particular research questions and are produced by different investigators working within different methodological traditions. No inference is made concerning what any individual contributor personally knows or believes about representational measurement. The endorsement probabilities used in the interrogation instead characterize the extent to which propositions are explicitly recognized, implicitly assumed or repeatedly reinforced across the publicly represented journal knowledge base.

This distinction is important because publication creates cumulative methodological effects. Practices repeatedly represented in a leading journal acquire visibility, legitimacy and familiarity. They are cited, taught, incorporated into subsequent research, reproduced in manufacturer submissions and consultancy, and encountered by researchers who subsequently contribute to HTA agencies, universities and professional organizations. Earlier interrogation of the journal identified this broader influence, noting that its publication environment has consequences extending to universities, research centers, professional organizations and reimbursement agencies.

The *Value in Health* knowledge base can therefore be characterized as a mature, internationally influential and highly quantitative body of health-economic and outcomes-research literature. It provides extensive evidence of how contemporary HTA arithmetic is constructed, applied, refined and communicated. The present interrogation asks the question that must precede assessment of that arithmetic: does this knowledge environment establish the measurement properties required to determine whether the numerical operations it routinely represents are scientifically admissible?

That is the boundary against which the 24 canonical propositions and their categorical endorsement probabilities are assessed.

CATEGORICAL PROBABILITIES

In the present application, the interrogation is tightly bounded. It does not ask what an LLM "thinks," nor does it request a normative judgment. Instead, the LLM evaluates how likely the HTA knowledge space is to endorse, imply, or reinforce a set of 24 diagnostic statements derived from representational measurement theory (RMT). Each statement is objectively TRUE or FALSE under RMT. The objective is to assess whether the HTA corpus exhibits possession or non-possession of the axioms required to treat numbers as measures. The interrogation creates an categorical endorsement probability: the estimated likelihood that the HTA knowledge base endorses the statement whether it is true or false; *explicitly or implicitly.*

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

This is why categorical probabilities are essential. Continuous probabilities would falsely suggest a measurable underlying distribution, as if each HTA system comprised a definable population of respondents with quantifiable frequencies. But large language models do not operate on that level. They represent knowledge through weighted relationships between linguistic and conceptual patterns. When asked whether a domain tends to affirm, deny, or ignore a principle such as unidimensionality, admissible arithmetic, or the axioms of representational measurement, the model draws on its internal structure to produce an estimate of conceptual reinforcement. The precision of that estimate must match the nature of the task. Categorical probabilities therefore provide a disciplined and interpretable way of capturing reinforcement strength while avoiding the illusion of statistical granularity.

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

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

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

The LLM computes these categorical probabilities from three sources:

1. **Structural content of HTA discourse**
   If the literature repeatedly uses ordinal utilities as interval measures, multiplies non-quantities, aggregates QALYs, or treats simulations as falsifiable, the model infers high reinforcement of these false statements.
2. **Conceptual visibility of measurement axioms**
   If ideas such as unidimensionality, dimensional homogeneity, scale-type integrity, or Rasch transformation rarely appear, or are contradicted by practice, the model assigns low endorsement probabilities to TRUE statements.
3. **The model's learned representation of domain stability**
   Where discourse is fragmented, contradictory, or conceptually hollow, the model avoids assigning high probabilities. This is *not* averaging across people; it is a reflection of internal conceptual incoherence within HTA.

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

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

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

INTERROGATION STATEMENTS

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

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

Measurement Theory & Scale Properties

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

Measurement Preconditions for Arithmetic

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

Rasch Measurement & Latent Traits

12. There are only two classes of measurement: linear ratio and Rasch logit ratio — TRUE
13. Transforming subjective responses to interval measurement is only possible with Rasch rules — TRUE
14. Summation of Likert question scores creates a ratio measure — FALSE

Properties of QALYs & Utilities

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

Falsifiability & Scientific Standards

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

Logit Fundamentals

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

Latent Trait Theory

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

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.

ASSESSMENT OF THE *VALUE IN HEALTH* RESULTS: A KNOWLEDGE BASE AT THE LIMIT OF REFORM

The interrogation of the *Value in Health* knowledge base identifies a structural inversion at the foundations of the quantitative framework the journal has helped develop, scrutinize and disseminate (Table 1). The significance of the results lies not simply in the low endorsement of individual measurement propositions or the high endorsement of propositions inconsistent with representational measurement. It lies in the relationship between them. The knowledge base strongly possesses the arithmetic of legacy health technology assessment while weakly possessing the scientific conditions that would make that arithmetic admissible

TABLE 1: ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS *VALUE IN HEALTH*

STATEMENT	RESPONSE 1=TRUE 0=FALSE	ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY	NORMALIZED LOGIT (IN RANGE +/- 2.50)
INTERVAL MEASURES LACK A TRUE ZERO	1	0.20	-1.40
MEASURES MUST BE UNIDIMENSIONAL	1	0.15	-1.75
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.20
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.85	+1.75
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.90	+2.20
EQ-5D-3L PREFERENCE ALGORITHMS CREATE INTERVAL MEASURES	0	0.90	+2.20
THE QALY IS A RATIO MEASURE	0	0.95	+2.50
TIME IS A RATIO MEASURE	1	0.95	+2.50

MEASUREMENT PRECEDES ARITHMETIC	1	0.10	-2.20
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.90	+2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.10	-2.20
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	0.95	+2.50
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.85	+1.75
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.15	-1.75
QALYS CAN BE AGGREGATED	0	0.95	+2.50
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.70	+0.85
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.90	+2.20
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.65	+0.60
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FOR ASSESSING THERAPY IMPACT FOR LATENT TRAITS	1	0.05	-2.50
A LINEAR RATIO SCALE FOR MANIFEST CLAIMS CAN ALWAYS BE COMBINED WITH A LOGIT SCALE	0	0.60	+0.40
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.25	-1.10
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

This distinction determines whether the problem can be addressed by methodological reform. A statistical method can be refined, a model recalibrated, a reporting standard strengthened and a sensitivity analysis expanded. Measurement inversion is different. It occurs before any of these activities begin. If the quantities entering the analytical framework have not first been demonstrated to possess the properties required for the arithmetic performed upon them, everything downstream rests upon an unresolved measurement failure.

The difficulty begins even before the QALY. Quantitative science does not begin with a score, utility, endpoint or model parameter. It begins with an attribute. The scientific question must first specify what property is to be measured. Only then can a measurement system be constructed and its scale properties established. The present 24-item interrogation provides little evidence that this attribute-first sequence organizes the *Value in Health* knowledge base. The proposition that measures must be unidimensional receives an endorsement probability of only 0.15 (−1.75 logits). Recognition that the outcome of interest for a latent trait is possession of that trait reaches only 0.25 (−1.10). Recognition of the Rasch logit ratio scale as the basis for assessing therapy impact for latent attributes is at the negative extreme, 0.05 (−2.50).

These findings are particularly serious for a journal concerned extensively with patient-reported outcomes, health status, quality of life, symptoms, functioning and preference measurement. Such concepts are not self-measuring. Pain, fatigue or physical function must first be defined as attributes. Their unidimensionality must be established. A measurement system must then demonstrate how possession of the attribute is located on an invariant quantitative continuum. Without these requirements, numerical assignment substitutes for measurement construction. The interrogation suggests precisely this danger. The false proposition that summations of subjective instrument responses are ratio measures is reinforced at 0.90 (+2.20), while the false proposition that summation of Likert question scores creates a ratio measure reaches 0.95 (+2.50). The requirements for constructing measures of latent attributes are effectively absent while numerical manipulation of subjective responses is strongly embedded.

This represents more than inadequate attention to psychometrics. It indicates a failure within the knowledge base to maintain the necessary distinction between an attribute, observations concerning that attribute, numerical scores assigned to those observations and a measure of possession of the attribute. Once that distinction is lost, statistical sophistication can be mistaken for measurement sophistication. Reliability, responsiveness, statistical association and conventional validation do not establish equal units, invariance, a non-arbitrary zero or the scale properties required for subsequent arithmetic.

The central result is the proposition that measurement precedes arithmetic. Its endorsement probability is only 0.10 (−2.20). Recognition that the axioms of representational measurement are required for arithmetic is also 0.10 (−2.20), while recognition that multiplication requires ratio measurement is again 0.10 (−2.20). These results identify the structural failure. The principles that should determine whether arithmetic may begin occupy the strongly negative region of the knowledge base profile.

This cannot be dismissed as an esoteric dispute over measurement terminology. Multiplication and division are central operations in the quantitative methods represented throughout *Value in Health*.

If measurement status does not determine admissible arithmetic, there is no effective scientific checkpoint between numerical representation and mathematical manipulation. The problem is not that the journal performs arithmetic badly. Much of the arithmetic may be performed with considerable technical sophistication. The problem is logically prior: the knowledge base does not establish whether much of that arithmetic should be performed at all.

The QALY makes the measurement inversion impossible to avoid. The construction QALY = U × T requires multiplication. The interrogation strongly recognizes that time is a ratio measure, with an endorsement probability of 0.95 (+2.50). Yet recognition that multiplication requires ratio measurement is only 0.10 (−2.20). At the same time, the knowledge base strongly reinforces the false proposition that the QALY itself is a ratio measure, at 0.95 (+2.50).

This produces an extraordinary epistemic configuration. The knowledge base recognizes the measurement status of time, weakly possesses the rule governing multiplication, and nevertheless strongly reinforces the quantitative product created by multiplying time by a utility whose required proportional measurement properties have not first been established. The remaining results demonstrate that this is not an isolated ambiguity. The false proposition that ratio measures can have negative values receives 0.90 (+2.20). The false proposition that the QALY is dimensionally homogeneous receives 0.85 (+1.75). The false proposition that QALYs can be aggregated receives 0.95 (+2.50).

The pattern is sequential. The measurement requirement for multiplication is weakly possessed; the resulting QALY is strongly possessed; the dimensional legitimacy of that QALY is strongly assumed; and its subsequent aggregation is strongly reinforced. The arithmetic has effectively become evidence for itself. Once QALYs are routinely calculated, their existence as numbers appears to validate their treatment as quantities. Once they are aggregated, differences are calculated. Once differences exist, incremental QALYs appear available for division into incremental costs. Each numerical operation supplies the input for the next. None of these downstream operations can establish the measurement properties that were required before the first multiplication occurred.

Recognition that time is a ratio measure cannot rescue the QALY. Establishing the measurement status of time establishes only one side of the multiplication. Utility must independently possess the properties required to operate as a proportional adjustment to time. The requirement for multiplication cannot confer those properties upon utility. They must be demonstrated before multiplication occurs. If they are absent, calculation of the numerical product cannot manufacture them.

The consequences extend directly to cost-effectiveness analysis. Recognition that claims for cost-effectiveness fail the axioms of representational measurement is only 0.15 (−1.75). If the QALY has not been established as a legitimate quantitative measure, subtraction cannot manufacture a legitimate incremental QALY. If the incremental QALY is inadmissible, dividing incremental costs by it cannot manufacture a scientifically meaningful cost-per-QALY ratio. The ICER therefore provides no second opportunity to correct the measurement problem. It compounds it.

Legacy HTA is often defended through the apparent sophistication of the complete economic evaluation. Models integrate clinical evidence, costs, probabilities, utilities, survival, extrapolation and uncertainty. Yet integration cannot correct the measurement status of the individual quantities entering the model. A model can combine numbers perfectly and still fail scientifically if those numbers do not support the operations performed upon them. Arithmetic cannot rescue a failure of measurement by adding further arithmetic.

The same principle applies to increasingly sophisticated decision modelling. Probabilistic sensitivity analysis, microsimulation, scenario analysis, structural uncertainty assessment and model validation operate downstream of the quantities admitted to the model. None can create equal units. None can establish a non-arbitrary zero. None can demonstrate unidimensionality. None can transform an ordinal representation into a ratio measure merely because it has been assigned a probability distribution and repeatedly sampled.

The interrogation exposes a related contradiction concerning falsifiability. Recognition that non-falsifiable claims should be rejected is comparatively strong at 0.70 (+0.85). Yet the false proposition that reference-case simulations generate falsifiable claims is reinforced at 0.90 (+2.20). The knowledge base therefore appears to recognize falsifiability in principle while simultaneously reinforcing a modelling environment in which conditional simulated outputs acquire the status of empirical claims. This mirrors the measurement inversion itself: a scientific principle can be recognized abstractly without functioning strongly enough to constrain established methodological practice.

The question then becomes whether the existing framework can be repaired. The answer depends upon where the failure occurs. If the problem were an incorrect parameter, inadequate sensitivity analysis, poor reporting standard or defective modelling technique, repair would be appropriate. The present interrogation locates the failure before these activities. It lies in the sequence by which numerical representations are admitted to arithmetic without measurement first functioning as the scientific gatekeeper.

Adding greater discussion of representational measurement to existing methodological guidance would therefore be insufficient if utilities, QALYs, ICERs and reference-case modelling continued unchanged. Measurement cannot be appended downstream to justify arithmetic already accepted upstream. The measurement requirement must determine which numerical operations are permitted before those operations are performed. The framework cannot preserve its foundational arithmetic while simultaneously accepting measurement principles that invalidate that arithmetic.

Either multiplication requires appropriate ratio measurement or it does not. Either a non-arbitrary zero is required for ratio measurement or it is not. Either dimensional homogeneity constrains legitimate quantitative operations or it does not. Either measurement precedes arithmetic or arithmetic is permitted retrospectively to determine what counts as measurement. These are not differences that can be accommodated through another reference case, reporting checklist or methodological refinement.

This places *Value in Health* in an unusually consequential position. The journal is not merely an observer of legacy HTA. It has been one of the principal international publication environments through which utilities, QALYs, cost-effectiveness analysis, preference measurement, patient-reported outcomes and decision modelling have been developed, reviewed and normalized. This history does not establish culpability, nor does the interrogation demonstrate that individual authors, reviewers or editors consciously rejected representational measurement. The unit of analysis is the cumulative knowledge environment. The finding is institutional rather than personal.

Institutional responsibility, however, does not disappear merely because the failure is systemic. Once measurement inversion has been exposed, continued publication of the same analytical framework without addressing its measurement foundations becomes scientifically different from its historical publication before the problem was explicitly identified. The question confronting the journal is no longer whether it should publish more papers about measurement. It is whether measurement will become a condition for accepting quantitative claims.

Will authors multiplying time by health-state values be required to demonstrate that those values possess the complementary ratio properties required for multiplication? Will claims based upon latent attributes be required to establish the attribute, unidimensionality and an appropriate measurement continuum? Will peer review distinguish scores from measures? Will model inputs be assessed for admissible arithmetic before reviewers consider the sophistication of the model constructed from them? Unless requirements of this kind become effective scientific constraints, nothing fundamental has changed.

The implications extend directly to ISPOR because *Value in Health* does not exist in methodological isolation. ISPOR Good Practices provide formal professional guidance while *Value in Health* provides its flagship peer-reviewed publication environment. If interrogation of both knowledge bases identifies the same measurement inversion, the problem spans two major mechanisms through which the methodological standards of health economics and outcomes research are created, legitimized and reproduced.

That convergence is more consequential than identifying measurement failure in a single paper, author group or national HTA agency. It suggests that professional methodological guidance and the publication environment have reinforced the same inherited quantitative framework without measurement functioning as an effective upstream constraint. International consensus provides no defense. If the same framework is taught, recommended, published, reviewed, cited and applied internationally without independent demonstration of its measurement foundations, widespread agreement demonstrates diffusion of the convention rather than validation of the measurement. What has been replicated is the methodology, not the measurement proof.

The implications for *Value in Health* are consequently impossible to avoid. Incremental reform is insufficient if reform means preserving the foundational arithmetic while leaving its measurement status unresolved. The scientifically relevant alternative is abandonment of the inverted analytical sequence. This does not mean abandoning quantitative therapy assessment, economic evidence or scientific publication. It means replacing arithmetic-first analysis with the required sequence of attribute, measurement, scale, admissible arithmetic, therapy-impact claim, empirical evaluation, replication and falsification.

For manifest attributes this requires appropriate linear ratio measurement. For latent attributes it requires a measurement system capable of establishing possession of a defined unidimensional attribute; within the framework applied here, this is provided by Rasch measurement. This is not another modelling preference competing with legacy HTA. It restores the measurement requirements that must be satisfied before modelling can legitimately begin.

The interrogation therefore places *Value in Health* at a scientific crossroads. Continuing to refine legacy HTA without resolving its measurement foundations would preserve precisely the inversion identified by these results. Accepting measurement as the gatekeeper, by contrast, requires reconsideration of methods whose arithmetic cannot satisfy the measurement conditions imposed upon them. There is no scientifically coherent middle position in which measurement principles are acknowledged but exempted whenever they threaten established methods.

The challenge to *Value in Health*, and by extension ISPOR, is therefore straightforward. Either the foundational quantities of legacy HTA can be demonstrated to possess the measurement properties required for the arithmetic performed upon them, or that arithmetic can no longer be defended merely because it is familiar, widely published, professionally endorsed or useful for decision making. If the required measurement demonstration cannot be provided, the consequence is not another refinement of legacy HTA. It is abandonment of the framework that permitted arithmetic to precede measurement.

THE PROGRAM CONTENT OF ISPOR EDUCATION

The measurement inversion identified in the *Value in Health* knowledge base acquires additional significance when considered alongside a separate interrogation of the educational program content of ISPOR [6] . This assessment addresses a different question from the 24-item knowledge-base interrogation. Rather than examining possession of canonical propositions concerning measurement and contemporary HTA, it asks whether the publicly available educational, training and competency content provided by ISPOR establishes the measurement requirements that should precede instruction in the quantitative methods of health economics, outcomes research and HTA.

The assessment used the frozen 37-topic framework previously applied to national and regional HTA program-content knowledge bases. Seventeen topics represent the requirements of representational measurement and 20 represent the methods and concepts characteristic of contemporary HEOR and HTA. The distinction between the two groups is fundamental. They are not parallel collections of methodological topics. The 17 representational-measurement topics establish the conditions governing whether the numerical operations employed in the 20 contemporary topics are scientifically admissible. Measurement is therefore logically prior to the methods whose arithmetic it governs.

The result of the ISPOR program-content interrogation was categorical: 0/17 representational-measurement topics compared with 20/20 contemporary HEOR and HTA topics. Publicly available ISPOR educational content provides extensive coverage of economic evaluation, utilities, QALYs, cost-effectiveness and cost-utility analysis, patient-reported outcomes, decision analysis, simulation modelling, uncertainty analysis, value-of-information analysis, budget impact and value assessment. Yet the interrogation identified none of the 17 representational-measurement topics required to establish whether the quantities entering these methods possess the measurement properties necessary for the arithmetic subsequently performed upon them.

The importance of the 0/17 result should not be understated by describing it as a curriculum gap or an imbalance in educational coverage. The missing topics are not optional additions to a comprehensive HEOR curriculum. They include the scientific requirements that determine what constitutes a measure, the properties associated with different scales of measurement, the admissible transformations associated with those scales, the requirements governing multiplication and division, dimensional homogeneity, unidimensionality and the different measurement pathways required for manifest and latent attributes. These requirements determine whether arithmetic can legitimately begin.

The contrast with the 20/20 result is therefore decisive. ISPOR educational content comprehensively represents the downstream analytical framework while the assessed representational-measurement foundation is entirely absent. Students and practitioners can encounter utilities, QALYs, ICERs, preference measurement, cost-effectiveness analysis and sophisticated decision modelling without first being required to establish why the numerical representations entering those methods are measures, what scales they possess or whether the arithmetic performed upon them is admissible. The educational sequence is inverted.

This result provides an important additional perspective on the measurement inversion identified in *Value in Health*. The journal interrogation demonstrates weak possession of the measurement requirements alongside strong possession of legacy HTA arithmetic. The program-content interrogation identifies a corresponding structure in the educational environment: the arithmetic and methods are comprehensively represented, while the measurement principles that should govern them are not represented among the 17 assessed topics. The two results were generated by different interrogation frameworks, but they point to the same underlying scientific problem.

The program-content result also helps explain how measurement inversion can persist within a mature and technically sophisticated professional environment. Researchers cannot be expected routinely to challenge the measurement foundations of methods if their professional education introduces those methods without first establishing the measurement standards against which they should be judged. A practitioner taught how to construct a QALY, calculate an ICER, estimate utilities, undertake probabilistic sensitivity analysis and develop a reference-case model may become highly proficient in those procedures while never being required to ask whether multiplication of utility by time is admissible in the first place. Technical competence can therefore increase while the logically prior measurement question remains unanswered.

This distinction is essential. The 20/20 result demonstrates substantial educational possession of contemporary HEOR and HTA. It says nothing, however, about whether those methods constitute admissible quantitative science. Competence in executing an arithmetic operation is different from demonstrating that the quantities involved permit that operation. A model can be correctly programmed, uncertainty correctly propagated and an ICER correctly calculated in a mathematical sense while the underlying quantitative claim remains scientifically inadmissible because the required measurement properties were never established.

The 0/17 result therefore closes the gate to the 20/20 framework. Downstream sophistication cannot compensate for an upstream failure of measurement. Adding more advanced modelling, more elaborate preference methods or improved computational techniques cannot supply the missing measurement properties because the failure occurs before these methods are reached. Measurement must establish what arithmetic is permitted; arithmetic cannot retrospectively establish that its inputs were measures.

Nor can representational measurement now be added to the ISPOR educational framework as another competency alongside economic evaluation, modelling, patient-reported outcomes or value assessment. To do so would preserve the inversion. Representational measurement is not one specialty among many within quantitative HTA. It provides the rules determining whether the quantitative operations employed by those specialties are admissible. Acceptance of measurement as the gatekeeper therefore requires abandonment of educational content wherever the methods being taught cannot satisfy the measurement requirements governing their arithmetic.

For therapy-impact assessment, the replacement begins with the attribute rather than the analytical technique. Manifest attributes require appropriate linear ratio measurement. Latent attributes require construction of a measure of possession of a defined unidimensional attribute through Rasch measurement. Only after measurement has been established and its scale properties demonstrated can the admissible arithmetic be determined. The sequence is the reverse of that represented by the 0/17 versus 20/20 result.

The program-content interrogation therefore adds an important third dimension to the assessment of ISPOR. Good Practices represent formal methodological endorsement. Value in Health represents the principal journal environment in which the methodology is developed, published and normalized. ISPOR education represents the mechanism through which methodological knowledge is transmitted to practitioners and subsequent generations of researchers. The evidence now extends across endorsement, publication and education.

This convergence makes it increasingly difficult to characterize measurement inversion as an isolated methodological oversight. The same underlying structure is represented across three distinct components of the ISPOR environment. The methods are endorsed, the methods are published and the methods are taught, while the representational-measurement requirements that should determine whether their arithmetic is scientifically admissible remain weakly possessed or absent.

The educational finding is therefore particularly consequential for any proposed abandonment of legacy HTA. ISPOR cannot abandon measurement inversion while continuing to teach the analytical framework through which measurement inversion is reproduced. Abandonment must include the mechanism of transmission. The 0/17 versus 20/20 result means that the educational architecture itself must be replaced by one in which measurement precedes arithmetic. Until that occurs, each new generation trained within the existing framework risks reproducing the same inversion that is evident in Good Practices and *Value in Health*.

CONCLUSION

The interrogation of the *Value in Health* knowledge base identifies pervasive measurement inversion. The arithmetic of legacy HTA is strongly represented while the measurement requirements that determine whether that arithmetic is admissible are weakly possessed or effectively absent. This is not an isolated journal phenomenon. Interrogations of other major HTA and health-economic journal knowledge bases identify essentially the same profile, while the separate interrogation of ISPOR Good Practices identifies the same underlying failure in the professional methodological environment. What has been replicated is the methodology, not the measurement proof.

The position of *Value in Health* is particularly consequential because it is ISPOR's flagship journal and a major international publication environment for utilities, QALYs, cost-effectiveness analysis, ICERs, preference measurement and decision modelling. The problem is not that these methods require further refinement. If their foundational arithmetic cannot satisfy the requirements of representational measurement, the measurement-inverted framework on which they depend must be abandoned. Greater modelling sophistication, probabilistic analysis, peer review, professional consensus and forty years of publication cannot manufacture measurement properties that were required before the arithmetic began.

The separate interrogation of ISPOR educational program content makes the institutional problem still more difficult to dismiss. Using the frozen 37-topic framework, the result was categorical: 0/17 representational-measurement topics compared with 20/20 contemporary HEOR and HTA topics. These are not equivalent dimensions of methodological competence. The 17 measurement topics constitute the scientific gatekeeper for the 20 downstream topics. Failure at 0/17 therefore cannot be compensated by success at 20/20. Instead, the result demonstrates complete program-content inversion: the downstream methods are comprehensively represented while the measurement requirements governing their admissibility are absent.

The evidence consequently extends across three distinct components of the ISPOR environment. Good Practices endorse the methodology, *Value in Health* publishes and normalizes it, and ISPOR education transmits it. All point toward the same underlying structure: arithmetic before measurement. Representational measurement cannot be added to this framework as another methodological competency because it determines whether the existing competencies are scientifically admissible. Where the required measurement properties cannot be demonstrated, the methods dependent upon them must be abandoned.

This leaves a question substantially larger than the future of *Value in Health*. ISPOR has exercised global influence in establishing and disseminating the methodological framework of contemporary HEOR and HTA. If that framework must now be abandoned because of pervasive measurement inversion, the organization has an extraordinary methodological legacy to unravel. Whether ISPOR has the scientific willingness and institutional capacity to manage that abandonment, reconstruct its publication and educational infrastructure around lawful measurement, and retain its claim to international scientific leadership is a question that now warrants examination in its own right.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

REFERENCES

[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

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

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

[4] Bond T, Zi Yan, Heene M, Applying the Rasch Model: Fundamental measurement in the human sciences (4th Ed.) . New York: Routledge 2021

[5] Langley P. Legacy HTA: ISPOR – The endorsement of measurement inversion in practice pattern standards. Logit Working Paper No 2000. https://maimonresearch.com/logit-working-paper-no-2000-august-2026/

[6] Langley P. ISPOR: Normal science and the absence of representational measurement in HTA education. Logit Working Paper No 3010. https://maimonresearch.com/logit-working-paper-no-3010-july-2026/