MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

THE MEASUREMENT FUTURE IN HEALTH TECHNOLGY ASSESSMENT

LEGACY HTA: ENDORSING MEASUREMENT INVERSION BY THE ISPOR CENTRAL AND EASTERN EUROPE (CEE) CONSORTIUM

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

LOGIT WORKING PAPER No 2003 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *Central and Eastern Europe (CEE) provides an important test of the scientific foundations of internationally transmitted health technology assessment (HTA). The ISPOR CEE Consortium currently comprises 18 chapters representing 15 countries and provides a regional environment for disseminating health economics and outcomes research (HEOR), including utilities, quality-adjusted life years (QALYs), cost-effectiveness analysis and decision modelling. The critical question is whether adoption of these methods has been accompanied by possession of the representational-measurement requirements necessary to justify their arithmetic.*

Methods: *The publicly accessible CEE HTA and HEOR knowledge environment was interrogated using a frozen set of 24 canonical statements derived from representational measurement theory and Rasch measurement principles. Each statement was classified TRUE or FALSE and assigned a categorical endorsement probability representing the extent to which the knowledge base supports the proposition explicitly or implicitly. Probabilities were transformed to normalized logits ranging from −2.50 to +2.50. The interrogation evaluates the conceptual structure of the knowledge base rather than the beliefs of individual practitioners.*

Results: *The CEE knowledge base demonstrates pronounced measurement inversion. The TRUE propositions* measurement precedes arithmetic*,* multiplication requires ratio measurement*, and* the axioms of representational measurement are required for arithmetic *each receive p = 0.15 (−1.75 logits)**.** In contrast, the FALSE propositions* the QALY is a ratio measure *and* the QALY is dimensionally homogeneous *each receive p = 0.90 (+2.20)**,** while* QALYs can be aggregated *reaches p = 0.95 (+2.50)**.** The inversion is equally pronounced for latent attributes. Rasch measurement propositions occupy the extreme negative boundary at p = 0.05 (−2.50)**,** whereas FALSE propositions treating summated subjective or Likert responses as ratio measures receive p = 0.90 (+2.20)**,** producing a 4.70-logit separation between lawful latent measurement and conventional score manipulation.*

*Conclusions**:** CEE strongly possesses the quantitative conclusions of legacy HTA while weakly possessing the measurement premises required to justify them. This should not be interpreted as a distinctive CEE failure. Rather, convergence with the wider ISPOR environment indicates successful international transmission of an inherited methodological architecture in which arithmetic precedes measurement. Methodological convergence is therefore not measurement validation. Reconstruction requires reversal of the sequence: attribute → measurement → scale properties → admissible arithmetic → empirically evaluable therapy-impact claim**.** Representational measurement cannot be added as another topic within legacy HTA; once restored as the scientific gatekeeper, methods that cannot satisfy its requirements must be abandoned.*

INTRODUCTION

Central and Eastern Europe (CEE) provides a particularly important setting in which to examine the international transmission of contemporary health technology assessment (HTA). The region encompasses countries with markedly different histories of healthcare financing, pharmaceutical reimbursement, economic evaluation and institutional development. Some have established formal HTA agencies and increasingly sophisticated reimbursement processes; others have developed these capabilities more recently. Across the region, however, health economics and outcomes research (HEOR), cost-effectiveness analysis, health-state utilities, quality-adjusted life years (QALYs), decision modelling and value assessment have become increasingly familiar components of the healthcare decision environment.

The ISPOR Central and Eastern Europe Consortium provides one of the professional structures through which these methods have been communicated, discussed and reinforced. As part of the wider ISPOR international environment, the Consortium connects researchers, academics, analysts, decision makers and other professionals with the methodological conventions of contemporary HEOR and HTA. Its significance therefore extends beyond professional networking. It forms part of an international knowledge infrastructure through which concepts of value, evidence, economic evaluation and quantitative decision making can become embedded within national research, educational and reimbursement environments.

This makes CEE an especially revealing test of the scientific foundations of internationally transmitted HTA. The relevant question is not whether the region has successfully adopted contemporary HEOR methodology. There is abundant evidence of methodological sophistication. The prior question is whether the quantitative framework being adopted satisfies the requirements of measurement necessary to give its arithmetic empirical meaning [1,2].

That question changes the terms of assessment. A methodology cannot be considered scientifically mature simply because it has become widely used, has been incorporated into reimbursement procedures or resembles methods employed in longer-established HTA systems. Quantitative legitimacy depends upon the properties of the quantities themselves. Before numbers are added, multiplied, divided, aggregated or entered into decision models, it must first be demonstrated that they are measures of defined attributes and that their measurement properties permit the arithmetic subsequently performed upon them.

The fundamental requirement is therefore straightforward: measurement precedes arithmetic. Scale properties determine admissible mathematical operations; mathematical operations cannot manufacture the scale properties they require. Multiplication requires the appropriate quantitative structure. Dimensional homogeneity must be established rather than assumed. Manifest and latent attributes require appropriate measurement pathways. Quantitative claims of therapy impact must ultimately be capable of empirical evaluation, replication and falsification. These requirements are logically prior to economic evaluation, modelling or uncertainty analysis because no degree of downstream analytical sophistication can repair a measurement failure occurring before the arithmetic begins.

This has immediate implications for the central constructions of legacy HTA. The QALY requires multiplication of a health-state value by time. Time presents no fundamental measurement difficulty because duration can satisfy the requirements for linear ratio measurement. The critical question is whether the health-state value possesses the measurement properties necessary to enter that multiplication. Those properties cannot be inferred from the existence of a preference algorithm, the assignment of numerical values to health states or the widespread institutional use of QALYs. They must be demonstrated independently. If that demonstration fails, the QALY does not acquire legitimacy merely because the multiplication can be executed computationally. Nor can subtraction to produce incremental QALYs or division to produce a cost-per-QALY ICER repair the original failure. Arithmetic cannot create measurement retrospectively.

The same requirement applies to latent attributes. Pain, fatigue, physical function, need fulfilment and similar attributes cannot be transformed into measures merely by assigning numbers to questionnaire responses and summing them. The attribute must first be defined and demonstrated to be unidimensional, and an invariant measurement continuum must be constructed. Without that prior measurement structure, a score remains a numerical representation rather than automatically becoming a measure capable of supporting whatever arithmetic an analyst wishes subsequently to perform. The necessary and sufficient solution is Rasch measurement [3] [4].

The CEE interrogation therefore asks a deliberately prior question. Does the publicly accessible HTA and HEOR knowledge environment associated with the ISPOR Central and Eastern Europe Consortium recognize the measurement requirements governing quantitative claims before endorsing the arithmetic of contemporary HTA? Or does it reproduce the reverse sequence: utilities, QALYs, aggregation, cost-effectiveness analysis and reference-case modelling strongly represented while representational measurement, ratio measurement, dimensional requirements and lawful latent measurement remain weakly possessed?

To address this question, the CEE knowledge base is interrogated using the same frozen set of 24 canonical statements applied across the wider assessment program. Each proposition has a stipulated status under the representational-measurement framework and is assigned a categorical endorsement probability representing the extent to which the knowledge environment possesses or reinforces it. These probabilities are transformed to normalized logits on a scale ranging from −2.50 to +2.50. Positive logits indicate stronger possession; negative logits indicate weak possession or conceptual absence. The purpose is not to estimate population frequencies or survey individual CEE researchers. It is to characterize the conceptual structure of the publicly accessible knowledge environment.

CEE is particularly valuable because it provides a test of the proposition that the measurement failure identified elsewhere is simply a peculiarity of mature Western HTA systems. If countries developing or expanding HTA more recently had reconstructed the discipline from its scientific foundations, a different pattern should be visible. Representational measurement should precede economic evaluation; ratio measurement should precede QALY multiplication; latent measurement should precede manipulation of patient-reported scores; and falsifiable empirical claims should precede hypothetical reference-case projections.

The results show that this has not occurred. Separating CEE from the wider European environment does not reveal a different foundational structure. Both display essentially the same 24-statement measurement inversion. This finding is important because it shifts the interpretation away from regional deficiency. CEE did not need independently to invent measurement inversion. It could inherit it through an international professional framework in which utilities, QALYs, cost-effectiveness analysis and modelling were already presented as established quantitative methods without representational measurement first operating as their scientific gatekeeper.

The CEE analysis is therefore not fundamentally an assessment of whether Central and Eastern European researchers have mastered contemporary HTA. It addresses the more disturbing possibility that they have mastered it successfully. If the international framework itself places arithmetic before measurement, faithful transmission will reproduce the same inversion wherever the methodology is adopted. Methodological convergence then ceases to provide evidence of scientific validation. It may instead provide evidence of successful intellectual transmission.

The question posed by the CEE interrogation is consequently not whether Central and Eastern Europe has caught up with contemporary HTA. It is whether what it has been encouraged to catch up with was scientifically defensible in the first place.

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 CENTRAL AND EASTERN EUROPE CONSORTIUM KNOWLEDGE BASE

The ISPOR Central and Eastern Europe (CEE) Consortium provides a regional professional framework for the development and dissemination of health economics and outcomes research (HEOR) and its application to healthcare decision making. ISPOR defines its consortia as regional groups comprising ISPOR chapters and individual members interested in advancing HEOR within their regions. The CEE Consortium consequently provides an institutional bridge between the wider ISPOR methodological environment and national academic, professional, health-policy and reimbursement communities across Central and Eastern Europe.

For the present interrogation, the CEE knowledge base is defined by the countries represented in ISPOR's current CEE Consortium chapter listing. The Consortium lists 18 chapters representing 15 distinct countries: Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Greece, Hungary, Israel, Poland, Romania, Russia, Slovakia, Slovenia, Turkey and Ukraine. Russia accounts for four separately identified chapters—the Russia Chapter, Russia-Far East Chapter, Russia HTA Chapter and Russia St. Petersburg Chapter—so the number of chapters should not be confused with the number of countries represented.

The geographical designation should also be understood as an ISPOR organizational category rather than a strict geographical definition of Central and Eastern Europe. The inclusion of Armenia, Israel and Turkey illustrates the breadth of the Consortium's institutional reach. Earlier ISPOR publications have also used broader CEE definitions for particular research purposes. For example, an ISPOR study of ICER thresholds examined 19 CEE countries, including Albania, Estonia, Latvia, Lithuania, Moldova, Montenegro, North Macedonia and Serbia, some of which are not currently listed as Consortium chapters. The present knowledge-base interrogation therefore uses the current ISPOR Consortium structure, rather than attempting to impose an independent geopolitical definition of CEE.

This is a substantial and heterogeneous knowledge environment. The countries differ in the maturity of their HTA systems, pharmaceutical reimbursement arrangements, economic circumstances, university infrastructures and experience with formal economic evaluation. ISPOR research itself demonstrates this diversity. An analysis of HTA requirements across 19 CEE countries found widespread use of budget-impact and cost-effectiveness analysis but differences in evidentiary requirements, local-data expectations and reimbursement procedures. More recent CEE work continues to examine differences in HTA processes and pharmaceutical access across the region.

The Consortium nevertheless provides a common professional environment. ISPOR describes it as a platform for information sharing and institutional collaboration in HEOR and health policy, supporting professional networking through scientific activities and global ISPOR initiatives. This creates a regional knowledge base extending beyond the output of any single national HTA agency or university. It encompasses the concepts, methodological assumptions and professional conventions represented through Consortium activities, national chapters, ISPOR educational and scientific activities, conference presentations, publications and the wider ISPOR HEOR environment to which regional participants are exposed.

The heterogeneity of the 15-country environment makes the CEE interrogation particularly important. If lawful measurement were independently embedded within national HTA development, substantial differences in foundational knowledge might be expected across such diverse systems. Conversely, if the countries have been exposed to a common international framework emphasizing utilities, QALYs, cost-effectiveness analysis, modelling and reimbursement decision making without representational measurement operating as a prior scientific gatekeeper, convergence toward the same measurement-inversion profile would be expected.

The CEE knowledge base is therefore not treated as though 15 countries constitute a single homogeneous intellectual community. Rather, it represents a shared regional HEOR environment within the ISPOR global structure. The interrogation asks what foundational measurement knowledge is possessed within that environment and, critically, whether the extensive transmission of contemporary HTA methodology has been accompanied by equivalent transmission of the measurement requirements necessary to establish that its quantitative operations are scientifically admissible.

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 ISPOR CENTRAL AND EASTERN EUROPE CONSORTIUM HTA KNOWLEDGE BASE

The 24-statement interrogation of the ISPOR Central and Eastern Europe (CEE) Consortium knowledge base reveals something more consequential than an educational omission or an imbalance between measurement theory and applied health economics. It identifies an internally structured knowledge system in which the conclusions of quantitative analysis are strongly possessed while the scientific premises required to justify those conclusions are weakly possessed or effectively absent (Table 1) The result is not simply measurement inversion in sequence. It is measurement inversion embedded in the epistemic architecture of the field.

TABLE 1: ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS CEE CONSORTIUM

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	0.20	-1.40
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.15	-1.75
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	0.90	+2.20
TIME IS A RATIO MEASURE	1	0.95	+2.50
MEASUREMENT PRECEDES ARITHMETIC	1	0.15	-1.75
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.90	+2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.15	-1.75
THERE ARE ONLY TWO CLASSES OF MEASUREMENT LINEAR RATIO AND RASCH LOGIT RATIO	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50

SUMMATION OF LIKERT QUESTION SCORES CREATES A RATIO MEASURE	0	0.90	+2.20
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.90	+2.20
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.15	-1.75
QALYS CAN BE AGGREGATED	0	0.95	+2.50
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.70	+0.85
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.85	+1.75
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.70	+0.85
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	0.20	-1.40
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

The pivotal result is the true proposition measurement precedes arithmetic, endorsed at only p = 0.15 (−1.75 logits). This is the gatekeeper proposition. If it were strongly possessed, every subsequent quantitative operation would be conditional upon prior demonstration of measurement. Instead, the CEE knowledge base strongly possesses the downstream arithmetic while weakly possessing the rule that determines whether that arithmetic is admissible. This single result goes a considerable way toward explaining the remainder of the profile.

The ratio-measurement results make the inversion explicit. Multiplication requires a ratio measure receives only p = 0.15 (−1.75 logits). Meeting the axioms of representational measurement is required for arithmetic is similarly weak at p = 0.15 (−1.75 logits). Yet the false proposition the QALY is a ratio measure is strongly possessed at p = 0.90 (+2.20 logits), while the false proposition QALYs can be aggregated reaches p = 0.95 (+2.50 logits). The contrast is not subtle.

The knowledge base has very weak possession of the requirement governing multiplication but extremely strong possession of the quantitative object produced by that multiplication.

This creates an extraordinary epistemic asymmetry. The CEE knowledge base effectively knows the answer before establishing that the calculation has empirical meaning. It accepts the QALY as a quantitative object without possessing the measurement rule required to construct it. It then accepts aggregation of those QALYs even more strongly. The methodological chain therefore runs backwards: \QALY accepted → aggregation accepted → cost-effectiveness accepted → measurement requirements largely absent. Lawful quantitative science requires the opposite sequence: attribute → measurement → scale properties → admissible arithmetic → quantitative claim.

CEE strongly possesses the first chain and weakly possesses the second. That is the essence of measurement inversion.

The QALY results are particularly damaging because the problem cannot be attributed to a general inability to recognize ratio measurement. The true proposition time is a ratio measure receives the maximum endorsement, $p = 0.95$ (+2.50 logits). CEE therefore recognizes a familiar ratio quantity perfectly well. What it does not strongly recognize is the requirement that the other quantity entering the multiplication must possess the properties necessary for the operation. The health-state utility is simply allowed to enter the calculation.

That distinction is crucial. Time cannot donate its ratio properties to utility. Multiplying a lawful ratio measure by a number whose required measurement properties have not been demonstrated does not create another lawful ratio measure. Yet the resulting QALY is strongly treated as one at $p = 0.90$ (+2.20 logits). This is not a minor technical imperfection in QALY methodology. It means that the quantitative status of the QALY is being inferred from the fact that the arithmetic is conventionally performed rather than demonstrated before the arithmetic is allowed.

The dimensional result reinforces the point. The false proposition the QALY is a dimensionally homogeneous measure receives $p = 0.90$ (+2.20 logits). Thus, the knowledge base not only strongly accepts the QALY as a ratio quantity; it also strongly accepts the dimensional legitimacy required for subsequent manipulation. The measurement proof is absent, but the downstream properties that would have followed from a successful proof are treated as though they had already been established.

This is epistemically backwards. In normal quantitative science, properties are demonstrated and operations follow. In the CEE profile, operations are institutionalized and the required properties are effectively presumed from their use.

The same pattern appears with subjective scores. The false proposition summations of subjective instrument responses are ratio measures receives $p = 0.90$ (+2.20 logits), as does summation of Likert questionnaire scores creates a ratio measure, again $p = 0.90$ (+2.20 logits). By contrast, measures must be unidimensional receives only $p = 0.20$ (−1.40 logits). The implication is striking. The knowledge environment strongly possesses the legitimacy of adding responses before it strongly possesses the requirement to establish that those responses represent a single measurable attribute.

This is more than a failure to teach unidimensionality. It indicates a conception of measurement in which generating a numerical score substitutes for demonstrating a quantitative structure. Once responses have been coded, weighted, summed or transformed, the numerical output acquires an authority that its empirical foundations have not earned. Arithmetic becomes the source of apparent measurement rather than the consequence of measurement.

The Rasch results show how deeply embedded this problem is for latent attributes. The proposition concerning the Rasch logit ratio pathway for therapy-impact assessment receives only p = 0.05 (−2.50 logits). The proposition linking Rasch measurement rules to the axioms of representational measurement is also p = 0.05 (−2.50 logits). The proposition concerning transformation of subjective responses through Rasch rules is again p = 0.05 (−2.50 logits). These values occupy the extreme negative boundary of the interrogation.

This produces an extraordinary contrast. The CEE knowledge base strongly possesses methods for assigning, summing and manipulating numbers attached to subjective responses at p = 0.90 (+2.20 logits) while possessing the lawful latent-measurement pathway at only p = 0.05 (−2.50 logits). The separation between these positions is 4.70 normalized logit units. The problem is therefore not that Rasch measurement receives somewhat less emphasis than conventional outcomes methods. The two occupy almost opposite extremes of the knowledge structure.

The same conclusion follows from the proposition that the outcome of interest for latent traits is possession of the trait, which receives only p = 0.20 (−1.40 logits). This exposes a deeper conceptual failure. Measurement of a latent attribute should establish where a person is located on a defined continuum and whether therapy changes possession of that attribute. Legacy HTA instead permits the analytical focus to migrate from possession of the attribute to scores, utilities, preferences and ultimately QALYs. The patient-level empirical attribute disappears behind increasingly elaborate numerical transformations.

Cost-effectiveness then inherits every preceding failure. The true proposition claims for cost-effectiveness fail the axioms of representational measurement receives only p = 0.15 (−1.75 logits). This is entirely predictable. Once the QALY has been granted quantitative legitimacy at p = 0.90 (+2.20) and its aggregation accepted at p = 0.95 (+2.50), the cost-per-QALY ICER appears to follow naturally. But it follows only arithmetically. Scientifically, nothing downstream can repair the missing measurement demonstration upstream. An inadmissible QALY does not become admissible when differences are calculated, and an inadmissible ΔQALY does not become a legitimate denominator when costs are divided by it.

Reference-case modelling takes the inversion one stage further. The false proposition reference-case simulations generate falsifiable claims receives p = 0.85 (+1.75 logits). Here the knowledge base moves beyond assuming measurement and begins treating simulated consequences of assumed measurement as though they possessed the epistemic status of empirical claims. The progression is revealing: measurement is not demonstrated; arithmetic proceeds nevertheless; the resulting quantities enter models; the models generate projected outcomes; and those projections are then accorded the status of quantitative evidence. Each downstream layer increases analytical sophistication while increasing distance from the unresolved measurement failure at the beginning.

This explains why technical sophistication cannot rescue the framework. More elaborate models, larger evidence networks, probabilistic sensitivity analysis, real-world data and increasingly powerful computation operate downstream of the measurement gate. If that gate has never been passed, sophistication merely performs more operations on quantities whose scientific status remains unresolved. Computational excellence and measurement legitimacy are independent properties.

The few positive lawful-measurement results do not alter this conclusion. Non-falsifiable claims should be rejected receives $p = 0.70$ (+0.85 logits), while the logit is the natural logarithm of the odds ratio also receives $p = 0.70$ (+0.85 logits). These results demonstrate fragments of relevant knowledge. But possession of isolated propositions is not possession of a coherent measurement framework. Knowing what a logit is while failing to possess the role of Rasch measurement at $p = 0.05$ (−2.50) does not establish latent measurement. Endorsing falsifiability abstractly while strongly treating reference-case simulations as falsifiable at $p = 0.85$ (+1.75) does not establish a falsificationist science.

The CEE profile is therefore characterized not merely by ignorance but by systematic epistemic displacement. The knowledge space is richest where it should be conditional and poorest where it should be foundational. QALYs, aggregation, preference algorithms and reference-case methods are strongly possessed. Measurement before arithmetic, ratio requirements, representational-measurement axioms and Rasch measurement are weakly possessed. The downstream apparatus has displaced the upstream science.

This is why the result cannot sensibly be described as an incomplete HTA framework. An incomplete framework could be repaired by adding missing material. CEE presents a more difficult problem because introducing lawful measurement as the gatekeeper would challenge methods already strongly embedded within the knowledge base. Teaching that multiplication requires ratio measurement at 0.15 (−1.75) is not simply filling a curricular gap when the QALY is simultaneously embedded at 0.90 (+2.20). Taking the former seriously requires reconsidering the latter. Introducing representational measurement therefore does not supplement legacy HTA; it destabilizes it.

The regional context makes the finding more important. Central and Eastern European countries differ substantially in the maturity of their HTA institutions and in the histories of their reimbursement systems. Yet separating CEE from wider Europe does not reveal a different foundational structure: both reproduce essentially the same measurement and program-content inversions. The reasonable interpretation is therefore not that CEE independently arrived at these errors. It inherited an international HEOR architecture in which they were already embedded.

That changes the meaning of methodological convergence. Similarity between CEE and established Western HTA environments cannot be treated as independent validation of the methods concerned. If both have inherited substantially the same assumptions from the same international professional environment, agreement demonstrates transmission, not validation.

Repetition of a methodology across jurisdictions is not replication of its measurement proof. As the wider analysis concludes, what has been replicated is the methodology, not the demonstration that its arithmetic is scientifically admissible.

The most disturbing feature of the CEE results is therefore their coherence. Measurement precedes arithmetic: 0.15/−1.75. Multiplication requires ratio measurement: 0.15/−1.75. Representational-measurement axioms are required for arithmetic: 0.15/−1.75. The QALY is a ratio measure: 0.90/+2.20. The QALY is dimensionally homogeneous: 0.90/+2.20. QALYs can be aggregated: 0.95/+2.50. These are not six unrelated observations. Together they describe the architecture of measurement inversion in almost ideal form. The CEE knowledge base strongly possesses what legacy HTA wants to calculate and weakly possesses the scientific rules that determine whether those calculations are permissible. That is the indictment.

The problem is not that CEE needs more sophisticated HTA. It already participates in a sophisticated international methodology. The problem is that sophistication has been constructed on the wrong side of the measurement gate. Adding further modelling, evidence synthesis, uncertainty analysis or computational refinement cannot correct that failure. The sequence itself must be reversed.

CEE does not need another refinement of legacy HTA. It needs to abandon the premise that repeated use, international acceptance and numerical sophistication can substitute for measurement demonstration. The reconstruction begins at the beginning: define the attribute, establish the measure, demonstrate its scale properties, determine the admissible arithmetic, and only then make the quantitative claim. Everything that cannot survive that sequence must be discarded.

TRANSFORMATION TO LAWFUL MEASUREMENT

Once measurement inversion has been demonstrated, there is only one scientifically defensible avenue forward: reconstruction from measurement. Legacy HTA cannot be repaired by improving utilities, refining QALYs, elaborating reference cases or constructing more sophisticated decision models. These activities occur downstream of the foundational failure. If the numerical representations entering the arithmetic have not first been established as measures possessing the properties required for that arithmetic, additional arithmetic cannot correct the problem. Nor can professional consensus, institutional acceptance or decades of application provide the missing measurement demonstration. The required transformation is therefore not reform of legacy HTA but its replacement by a framework in which measurement precedes arithmetic.

That transformation is already operational. Maimon Research has developed a nine-unit program for rebuilding HTA around the requirements of representational measurement, designed both for independent study and as the foundation for a three-credit graduate course [5] . The program begins where quantitative science must begin: with the definition of the attribute and the requirements that must be satisfied before numerical operations are permitted. It then distinguishes the two measurement pathways required for therapy-impact claims: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes.

Across the nine units, participants move from measurement inversion and the foundations of representational measurement through scale properties, admissible transformations and arithmetic, dimensional homogeneity, manifest and latent attributes, construction of Rasch measurement systems, possession and change, and the requirements for empirically evaluable therapy-impact claims. The sequence culminates in application, replication and falsification.

The transformation therefore requires neither another generation of methodological debate nor permission from an HTA agency or professional organization. The scientific sequence is straightforward: attribute → measurement → scale → admissible arithmetic → therapy-impact claim → empirical evaluation → replication → falsification. The program is available now. The remaining task is application.

CONCLUSION

The interrogation of the ISPOR Central and Eastern Europe (CEE) Consortium HTA knowledge base identifies a coherent and deeply embedded pattern of measurement inversion. The importance of the finding lies not simply in the weak representation of measurement theory, but in the extraordinary asymmetry between the scientific requirements that should govern quantitative analysis and the numerical conventions that legacy HTA routinely applies. The knowledge base strongly possesses the arithmetic while weakly possessing the conditions that would make that arithmetic scientifically admissible.

The contrast is stark. Measurement precedes arithmetic receives an endorsement probability of only 0.15 (−1.75 logits). Multiplication requires ratio measurement is likewise 0.15 (−1.75), as is the requirement that the axioms of representational measurement govern arithmetic, 0.15 (−1.75). Yet the false proposition that the QALY is a ratio measure receives 0.90 (+2.20), that the QALY is dimensionally homogeneous receives 0.90 (+2.20), and that QALYs can be aggregated reaches 0.95 (+2.50). The knowledge base therefore strongly possesses the quantitative conclusions of legacy HTA while scarcely possessing the measurement premises necessary to justify them.

This is measurement inversion reduced to numbers. The QALY is accepted before the admissibility of its defining multiplication has been established. Aggregation is accepted before the quantitative status of what is being aggregated has been demonstrated. Cost-effectiveness analysis then builds upon the resulting quantities, and reference-case simulation carries them into modeled futures. Each additional analytical layer increases methodological sophistication while moving further downstream from an unresolved measurement failure. Arithmetic has not followed measurement; arithmetic has substituted for it.

The latent-attribute findings deepen the indictment. Propositions identifying the Rasch measurement pathway occupy the extreme negative end of the profile at 0.05 (−2.50 logits), while false propositions treating summed subjective responses and Likert questionnaire scores as ratio measures receive 0.90 (+2.20). The resulting 4.70-logit separation captures an extraordinary reversal of scientific priorities. The knowledge environment strongly possesses methods for manipulating scores while effectively failing to possess the measurement framework required to establish a legitimate latent quantitative attribute in the first place.

These findings should not be interpreted as evidence of a distinctive Central and Eastern European scientific failure. CEE contains countries with different histories of HTA, reimbursement arrangements, academic infrastructures and levels of institutional maturity. Yet separating CEE from the wider European environment does not reveal a fundamentally different epistemic structure. The same measurement inversion persists. The more plausible interpretation is international transmission: CEE has inherited and reproduced an established HEOR and HTA framework in which representational measurement was never installed as the gatekeeper for arithmetic.

This distinction removes one possible defense. International methodological convergence cannot validate the inherited framework. If CEE, Western Europe and other regions employ substantially the same methods because they have inherited substantially the same professional conventions, their agreement is not independent scientific confirmation. It demonstrates successful dissemination. What has been replicated is the methodology, not the measurement proof.

Nor can the problem now be resolved by adding representational measurement as another topic within the existing framework. Measurement is not an optional competency alongside utilities, QALYs, economic evaluation and modelling. It determines whether those downstream operations are admissible. Once measurement is restored to its logically prior position, methods that cannot pass the measurement gate cannot simply be retained because they are familiar, professionally accepted or institutionally convenient.

The implication for the CEE Consortium, its national chapters, universities, HTA agencies and individual researchers is therefore unavoidable. Inheritance explains how measurement inversion became established; it cannot justify its continued transmission once the failure has been exposed. Another generation of researchers cannot be taught that the QALY is a legitimate quantitative construct while the requirement governing its multiplication remains at 0.15 (−1.75 logits). Nor can international acceptance transform an inadmissible operation into an admissible one.

The required transition is consequently not refinement but reconstruction. HTA must begin with the attribute, establish its measurement requirements, distinguish manifest from latent attributes, demonstrate the appropriate measurement structure, determine the arithmetic permitted by that structure, and only then advance quantitative therapy-impact claims capable of empirical evaluation, replication and falsification. Methods that survive this sequence can remain. Those that do not must be abandoned.

CEE therefore faces a straightforward scientific choice. It can continue reproducing an internationally inherited framework in which the arithmetic is strongly possessed and its measurement foundations largely absent, or it can break with that inheritance and reconstruct HTA from measurement forward. No additional international consensus is required and no professional organization can waive the requirements involved. The measurement failure has been exposed. Continuing to reproduce it is now a choice

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. New York: Routledge, 2021

[5] Maimon Research, Transformation Program. https://maimonresearch.com/hta-reconstruction-program-and-fees/