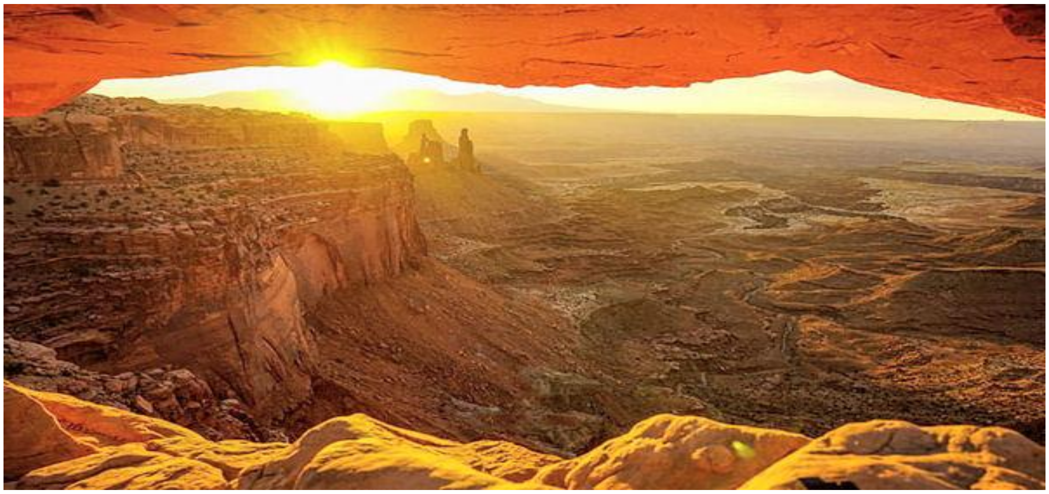
MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION



REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

UNITED KINGDOM: DECONSTRUCTING THE EPISTEMIC KNOWLEDGE BASE OF THE EQ-5D-3L AND EQ-5D-5L INSTRUMENTS

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

LOGIT WORKING PAPER No 43 FEBRUARY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper examines the EQ-5D-3L and EQ-5D-5L instruments not as descriptive questionnaires or psychometric tools, but as epistemic objects embedded within the United Kingdom health technology assessment (HTA) system. In HTA practice, instruments are commonly treated as neutral measurement devices whose numerical properties are fixed at development and merely applied thereafter. This assumption is rejected. Instruments acquire numerical authority through use. Numbers become treated as quantities when professional communities routinely subject them to arithmetic operations, regardless of whether the axioms required for measurement are satisfied.

The EQ-5D occupies a distinctive position within HTA because it functions as the primary descriptive and valuation foundation of the reference-case framework. Its outputs are routinely analyzed as outcomes, transformed into utilities, multiplied by time to generate quality-adjusted life years, and aggregated for policy decision-making. Over time, this repeated application has conferred an appearance of quantitative legitimacy that is rarely examined at the level of measurement theory. The transition from EQ-5D-3L to EQ-5D-5L is commonly interpreted as methodological refinement, yet whether such expansion alters the epistemic status of the instrument remains unexamined.

To address this gap, the paper defines the EQ-5D knowledge base as an epistemic corpus encompassing developers, users, national value sets, methodological guidance, educational materials, and analytic infrastructure. This corpus is interrogated using a reduced canonical diagnostic grounded in representational measurement theory and Rasch principles. The diagnostic evaluates whether the knowledge base recognizes the axioms required for quantitative measurement, including unidimensionality, invariant transformation, scale-type coherence, and arithmetic admissibility. Endorsement probabilities are classified and transformed into normalized logits to reveal structural patterns of epistemic reinforcement or absence.

The results display a coherent and invariant profile. Core measurement axioms receive no positive reinforcement, while assumptions enabling arithmetic treatment of utilities remain normalized. Rasch requirements for latent-trait measurement are entirely absent, consistent with the multiattribute, preference-based architecture of the EQ-5D system. The expansion from three to five response levels does not alter these foundational commitments. The knowledge base exhibits non-possession of measurement principles rather than misunderstanding or dispute.

The analysis demonstrates that the EQ-5D functions as a valuation framework embedded within an epistemic system that treats valuation as measurement. Because the reference-case framework depends upon EQ-5D outputs, non-measurement is inherited rather than corrected at the policy level. The paper concludes that refinement of descriptive detail cannot resolve this condition and that restoring measurement as a precondition for arithmetic is necessary if HTA is o generate quantitative knowledge rather than numerical appearance. EQ-5D and similar multiattribute instruments should be abandoned for other than purely descriptive purposes.

I. INTRODUCTION: THE USER EPISTEMIC SYSTEM

In health technology assessment, instruments are commonly treated as neutral devices that exist independently of the systems in which they are used. Once developed, they are assumed to generate numerical outputs whose meaning is fixed, transportable, and stable across applications. Under this view, the analytical task lies not in questioning whether numbers measure anything, but in deciding how best to deploy them. This assumption is deeply misleading. Instruments do not acquire numerical legitimacy through construction alone. They acquire it through use.

Numbers become treated as quantities only when a community repeatedly subjects them to arithmetic operations and accepts the results as meaningful. That acceptance is rarely explicit. It emerges through routine practice, institutional endorsement, educational transmission, and methodological repetition. Over time, what begins as assumption becomes belief, and belief becomes norm. Once embedded in this way, the question of whether an instrument truly measures anything at all no longer appears as a scientific problem. It becomes epistemically invisible.

The EQ-5D-3L and EQ-5D-5L instruments provide a particularly revealing illustration of this process [1,2] Originally developed as descriptive systems intended to classify health states, the EQ-5D instruments now occupy a central position within the United Kingdom's health technology assessment environment. Their numerical outputs are treated as utilities, incorporated into economic models, multiplied by time to generate quality-adjusted life years, and aggregated to support population-level decision-making. These operations are not occasional or experimental. They are routine. Through repetition, numerical treatment has come to substitute for measurement justification; meeting the axioms of representational measurement [3,4] .

This paper begins from the premise that numerical authority is socially conferred. Numbers do not become measures because they appear precise, because they are produced algorithmically, or because they are widely used. They become measures only when they satisfy the axioms that govern representation of empirical attributes. Those axioms specify when numbers can legitimately stand in for quantities and when arithmetic operations are permissible. Where such axioms are not recognized as governing constraints, numerical manipulation may proceed, but measurement has not occurred. Numerical storytelling dominates.

The object of analysis is therefore not the EQ-5D instrument in isolation, nor the intentions of its developers, nor the technical details of its valuation protocols. The object of analysis is the epistemic system within which EQ-5D outputs function as if they were quantitative measures. This paper refers to that system as the user epistemic system; it does not confer measurement legitimacy.

The user epistemic system consists of the collective practices through which EQ-5D values are interpreted as meaningful numbers. It includes researchers who analyze EQ-5D data, reviewers who evaluate manuscripts, editors who publish results, health technology assessment agencies that accept EQ-5D utilities as inputs, educators who train analysts in their application, and software environments that embed scoring algorithms. Together, these actors form a distributed but coherent knowledge base. No single participant determines its structure, yet each reinforces it.

Importantly, this system does not operate through explicit agreement about measurement theory. There is no formal declaration that EQ-5D utilities satisfy the axioms required for quantification. Instead, authority arises through practice. Means are reported. Differences are compared. Utilities are multiplied by time. Each step appears innocuous in isolation. Collectively, they construct a powerful presumption: if the numbers are used as quantities, they must be quantities. Use becomes evidence; repetition becomes validation.

Within such a system, epistemic responsibility is diffuse. Developers may point to widespread adoption. Users may point to methodological guidance. Agencies may point to precedent. Educators may point to accepted curricula. Each component defers foundational justification to another. The result is epistemic closure: numerical practice persists without ever encountering the conditions that would authorize or prohibit it.

This distinction between ignorance and non-possession is critical. The issue is not that users of the EQ-5D fail to understand measurement theory. It is that measurement theory does not function as a governing authority within the system. Where axioms are not recognized, they cannot constrain practice. Arithmetic proceeds not because rules are violated, but because the rules are absent.

The transition from EQ-5D-3L to EQ-5D-5L exemplifies this dynamic. The expansion of response levels is commonly interpreted as methodological progress, improving sensitivity and discrimination. Yet increased descriptive granularity does not address the prior question of whether the resulting numbers represent a measurable attribute. Refinement operates entirely within an already accepted numerical framework. It presupposes measurement rather than establishing it.

This paper therefore does not ask whether EQ-5D instruments are useful, convenient, or widely adopted. It asks a more fundamental question: does the knowledge base that authorizes their numerical use contain the axioms required for measurement? By treating EQ-5D as an epistemic object embedded within a user system rather than as a technical artifact, the analysis shifts attention from instrument performance to the conditions that make numerical authority possible.

The sections that follow proceed accordingly. Section II defines the EQ-5D knowledge base as an epistemic corpus encompassing developers and users alike. Section III describes the interrogation of this corpus using a reduced canonical diagnostic grounded in representational measurement theory and Rasch principles. Section IV presents the results of that interrogation. Section V considers the implications for instrument legitimacy and the broader structure of health technology assessment. Through this approach, the paper seeks to render visible what routine use has obscured: that numerical authority in HTA is not discovered through measurement, but constructed through belief.

II. THE EQ-5D INSTRUMENT KNOWLEDGE BASE

Having established that numerical authority arises through use rather than construction alone, the next task is to define what constitutes the knowledge base of the EQ-5D system. This cannot be limited to the original development papers of the EuroQol Group, nor can it be confined to formal descriptions of the instrument. Once released into applied domains, an instrument becomes embedded within a far broader epistemic environment. Its authority is sustained not by its design history, but by the network of texts, practices, institutions, and routines that treat its outputs as quantities. The EQ-5D knowledge base must therefore be understood as an epistemic corpus.

This corpus includes the foundational publications introducing the EQ-5D-3L and later the EQ-5D-5L. These texts define the descriptive architecture of the instrument: a multiattribute classification system covering mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each expressed through ordered response levels. From the outset, the instrument was framed not as a direct measure of health, but as a system for describing health states to be valued through preference elicitation. This framing is crucial. It establishes valuation, rather than measurement, as the conceptual foundation of the instrument.

However, the authority of the EQ-5D does not persist because these development papers are repeatedly interrogated. It persists because subsequent users treat the resulting numerical outputs as if they were quantitative measures. The dominant component of the knowledge base therefore lies in applied research. Thousands of clinical trials, observational studies, and economic evaluations report EQ-5D values as outcomes. Means are calculated, changes interpreted, and group differences compared. These operations are typically presented without discussion of scale type, unidimensionality, or invariance. Yet the absence of such discussion is itself epistemically powerful. It signals that justification is unnecessary.

Health technology assessment agencies constitute a second and particularly influential layer of the EQ-5D knowledge base. Within the United Kingdom, NICE guidance explicitly endorses EQ-5D as the preferred instrument for estimating health-related quality of life. This endorsement confers institutional authority. It does not arise from demonstration that EQ-5D values satisfy the axioms of measurement, but from their consistency with the reference-case framework. Once incorporated into official guidance, the instrument’s numerical status becomes administratively secured rather than theoretically established.

Methodological documents further reinforce this authority. Reference-case manuals, submission templates, and technical support documents routinely specify EQ-5D utilities as required or preferred inputs. These texts treat utilities as interchangeable numerical entities, abstracted from their descriptive origins. The instrument becomes a standardized component of analytic workflow rather than an object of epistemic scrutiny. At this stage, the question of what kind of numbers EQ-5D produces is displaced by the assumption that numbers are required.

Education plays a central role in reproducing this knowledge base. In graduate training programs and professional short courses, students are taught how to apply EQ-5D utilities in modeling exercises. They learn to compute quality-adjusted life years, to compare incremental cost-effectiveness ratios, and to interpret numerical thresholds. Rarely are they taught to interrogate whether the utilities themselves possess the properties required for arithmetic. By the time analysts enter practice, numerical legitimacy has already been internalized. The instrument is encountered not as a theoretical proposition, but as a given.

The epistemic reach of the EQ-5D knowledge base extends further through analytic infrastructure. Software packages, economic models, and spreadsheet templates embed EQ-5D scoring algorithms and national value sets. Once encoded, assumptions become invisible. Users interact with outputs without encountering the conceptual premises that authorize their numerical treatment. In this way, epistemic commitment is no longer expressed through argument or citation, but through automation.

The introduction of the EQ-5D-5L illustrates how the knowledge base absorbs modification without altering its foundations. The expansion from three to five response levels is presented as improved sensitivity and reduced ceiling effects. Yet this refinement does not alter the underlying epistemic architecture. The instrument remains multiattribute. Valuation remains preference-based. Utilities continue to allow negative values. Arithmetic compatibility with the QALY remains assumed rather than demonstrated. The proliferation of versions therefore signals elaboration within a fixed belief system rather than epistemic change.

Crucially, the EQ-5D knowledge base is not unified by explicit theoretical agreement. There is no authoritative text asserting that EQ-5D utilities satisfy representational measurement axioms. Instead, unity emerges through coordinated silence. Measurement theory is not debated because it is not invoked. Scale properties are not defended because they are not questioned. The absence of foundational discourse functions as a stabilizing mechanism.

This distributed structure explains the resilience of the EQ-5D system. Developers can point to widespread use. Users can point to agency guidance. Agencies can point to precedent. Educators can point to standard curricula. Each component defers epistemic responsibility to another. The result is a closed loop in which numerical authority circulates without ever encountering measurement constraints.

Defining the EQ-5D knowledge base in this way is essential for the analysis that follows. The purpose of interrogation is not to assess individual publications or intentions, but to determine whether the epistemic environment as a whole recognizes measurement axioms as governing rules. Only by treating the instrument as embedded within this broader corpus can its numerical status be meaningfully evaluated.

III. INTERROGATING THE EQ-5D KNOWLEDGE BASE

If the numerical authority of the EQ-5D system derives from a distributed epistemic environment rather than from instrument construction alone, then evaluating the instrument requires a method capable of interrogating that environment. The object of analysis is not whether individual studies apply the instrument correctly, nor whether its developers intended to create a measure. The relevant question is whether the knowledge base that authorizes numerical use recognizes the principles that determine when numbers can meaningfully represent quantities.

Interrogation, as used in this paper, refers to the systematic probing of conceptual reinforcement within a defined corpus. It does not seek beliefs, intentions, or opinions, nor does it assess methodological competence or good faith. Rather, it examines what the knowledge base articulates, normalizes, or excludes. The purpose is to determine which principles function as operative constraints on numerical interpretation and which are epistemically absent.

The theoretical foundation for this interrogation is representational measurement theory. Under this framework, measurement is not the assignment of numbers per se, but the construction of numerical representations that preserve empirically testable relational structure. Scale type determines permissible arithmetic operations. Ordinal scales support ordering only. Interval scales permit addition and subtraction but lack a true zero. Ratio scales alone permit multiplication, division, and meaningful comparison of magnitudes. These distinctions are not methodological conventions; they are logical preconditions for quantification.

Where these conditions are not satisfied, arithmetic operations are undefined regardless of how routinely they are performed in practice. Numerical appearance cannot substitute for representational validity.

These requirements are especially stringent when the attribute of interest is latent. Constructs such as health-related quality of life are not directly observable and must be inferred from response patterns. In such circumstances, ordinal observations cannot become quantitative without transformation through a model capable of producing invariant units. Rasch measurement is included in the interrogation framework not as a preferred methodology, but because it is the only model consistent with the axioms of representational measurement for latent attributes. Its role here is diagnostic: it defines the boundary separating ordinal scoring from measurement.

The interrogation therefore draws upon a canonical set of propositions derived from representational measurement theory and Rasch principles. Each proposition expresses either a necessary condition for measurement or a known impossibility when those conditions are violated. These propositions do not reflect theoretical preference. They specify the logical architecture required for numerical representation to be meaningful.

For purposes of instrument-level evaluation, a reduced canonical diagnostic is employed. This diagnostic is distinct from the twenty-four-item framework used in system-level assessments. The reduction is principled rather than procedural. Instruments do not authorize aggregation across populations, time-based multiplication, or cost-effectiveness modeling. Their epistemic responsibility lies upstream. They generate numerical outputs that may later be subjected to arithmetic by other actors. Whether such arithmetic is legitimate depends entirely on whether the instrument outputs qualify as quantities in the first place.

The reduced diagnostic therefore interrogates what may be termed the pre-arithmetic boundary. It examines whether the knowledge base recognizes the conditions under which numerical outputs could, in principle, function as measures before any modeling or policy application occurs. If measurement is not established at this boundary, no subsequent quantitative operation can be logically licensed.

Item selection follows this boundary logic. Statements are retained only if they test necessary conditions for measurement itself, including unidimensionality, scale type, invariance, the existence of a true zero, and the precedence of measurement over arithmetic. Statements that presuppose the existence of quantities — such as aggregation, time multiplication, or composite outcome construction — are excluded, as their validity depends entirely on whether measurement has already occurred.

Within the reduced set, propositions are conceptually classified according to the epistemic condition they test. One group addresses foundational axioms of representational measurement governing scale structure and arithmetic permission. A second group addresses latent attribute measurement and the necessity of invariant transformation. A third group interrogates valuation-based scoring and the substitution of preference for measurement. Together, these categories capture the principal mechanisms through which numerical form may be mistaken for quantitative meaning.

Interrogation does not assess whether these propositions are explicitly endorsed or denied in published texts. It assesses whether they function as operative constraints within the knowledge base. A principle may never be rejected and yet remain epistemically absent. Where axioms are not invoked, taught, or used to adjudicate claims, they do not exist in functional terms.

For this reason, endorsement probabilities are interpreted as indicators of possession rather than belief. Low endorsement does not imply disagreement. It indicates that the principle does not operate as a governing rule within the epistemic environment. Uniform endorsement profiles should therefore not be interpreted as confusion or inconsistency. They represent the expected signature of structural non-possession.

The interrogation thus functions as a diagnostic instrument. It does not adjudicate competence, intent, or methodological sophistication. Its purpose is narrower and more fundamental: to determine whether the epistemic environment in which the EQ-5D operates contains the conceptual conditions required for measurement. Where those conditions are absent, numerical practice may persist, but it does so without representational authorization.

Applied to the EQ-5D system, this framework allows the analysis to move beyond debates over descriptive sensitivity, valuation protocol, or empirical fit. Such debates presuppose that the instrument already produces quantities. The present interrogation suspends that presupposition. It asks instead whether the knowledge base that sustains the instrument recognizes the axioms required for quantitative representation at all.

The following section presents the results of this interrogation. It reports the endorsement profile for the reduced canonical statements and examines the internal coherence of the resulting configuration. Interpretation focuses not on individual propositions in isolation, but on the structure of the profile as a whole, as it is this structure that reveals the epistemic status of the EQ-5D system.

IV RESULTS AND DISCUSSION

As detailed in Table 1 the reduced canonical diagnostic applied to the EQ-5D-3L and EQ-5D-5L knowledge base yields a profile that is neither ambiguous nor transitional. Across the fourteen propositions, endorsement probabilities collapse toward the lower bound, producing uniformly negative normalized logits. The distribution exhibits internal coherence rather than noise: a small subset of foundational axioms registers weak-to-near-absent reinforcement, while the remainder collapses to the absolute floor. No proposition approaches neutrality, and none is positively reinforced. This is not a pattern of mixed commitments or partial compliance with measurement requirements. It is the signature of an epistemic environment in which the axioms of representational measurement do not function as admissibility conditions for numerical claims.

TABLE 1

REDUCED ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS EQ-5D INSTRUMENTS UK

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	0.10	-2.20
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.20
MEASUREMENT PRECEDES ARITHMETIC	1	0.10	-2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FO ASSESSING LATENT TRAIT IMPACT	1	0.05	-2.50
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.05	-2.50
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50
PREFERENCE BASED UTILITIES CREATE INTERVAL MEASURES	0	0.05	-2.50
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.05	-2.50
MULTIATTRIBUTE HEALTH STATE CLASSIFICATGIONS ARE UNIDIMENSIONAL	0	0.05	-2.50
ORDINAL HEALTH STATE DESCRIPTIONSCAN BE	0	0.05	-2.50

TRANSFORMED INTO QUANTITATIVE MAGNITUDE THROUGH PREFERENCE WEIGHTING			
PREFERENCE ALGORITHM SCORING PRODUCES MEASUREMENT VALID NUMERICAL QUANTITIES	0	0.05	-2.50

The first and least negative result concerns the proposition that interval measures lack a true zero (p = 0.20; −1.40). This placement is instructive because the EQ-5D environment frequently references the conventional "dead = 0" anchoring and routinely acknowledges the existence of "states worse than dead." These features, however, do not translate into a disciplined recognition of what a true zero is and why it matters. The diagnostic indicates that while the language of anchoring appears in the corpus, the concept does not operate as a governing constraint. In effect, the EQ-5D knowledge base treats zero as an administrative reference point rather than as an empirically grounded null of the attribute. A true zero is not a matter of convenience. It is the defining condition that separates ratio measurement from all other scale types because it represents the absence of the attribute being measured. Without a true zero, multiplication and ratio interpretation are undefined. The weak reinforcement here signals that this elementary requirement does not constrain how EQ-5D utilities are interpreted or used.

The problem becomes decisive when the diagnostic turns to unidimensionality. The proposition that measures must be unidimensional (p = 0.10; −2.20) collapses close to the floor, reflecting near-total non-possession of the principle as a governing rule. This matters because the EQ-5D is multiattribute by design. Whether in the 3L or 5L form, it describes health in terms of distinct domains, mobility, self-care, usual activities, pain/discomfort, anxiety/depression, each with ordered response levels. These domains do not constitute manifestations of a single empirical attribute with an established unit. They are heterogeneous descriptors. The diagnostic result indicates that the epistemic system does not treat this heterogeneity as a measurement barrier. The multiattribute structure is normalized and then treated as if it could yield a single magnitude. That normalization is precisely the epistemic move that the diagnostic is designed to expose. Measurement is not achieved by combining multiple domains under a single label; it requires demonstration that variation is along one attribute, preserving empirical structure under an admissible mapping.

The same near-floor outcome appears for the proposition that multiplication requires a ratio measure (p = 0.10; −2.20) and for the proposition that measurement precedes arithmetic (p = 0.10; −2.20). These two items jointly identify the foundational inversion that characterizes the EQ-5D system as used in HTA. In a measurement-governed discipline, arithmetic operations are conditional upon the scale properties of the numbers. In the EQ-5D environment, arithmetic is treated as a methodological necessity that defines the meaning of the numbers. Utilities are multiplied by time because the QALY framework requires multiplication, not because the utility numbers have been shown to support multiplication. The same is true of averaging, differencing, and aggregating. The instrument outputs are treated as quantities because they are used as if they were quantities. The diagnostic confirms that the knowledge base does not recognize the logical dependency in the correct direction. Measurement does not govern arithmetic; arithmetic governs what is treated as measurement.

This inversion is sealed by the proposition that meeting the axioms of representational measurement is required for arithmetic (p = 0.05; −2.50). Here the diagnostic reaches the absolute floor. The implication is not merely that the EQ-5D corpus does not explicitly articulate representational measurement axioms, but that these axioms do not function as admissibility criteria in practice. Where axioms are not part of disciplinary grammar, they cannot constrain routine numerical operations. This is the structural definition of non-possession. It is not that the system debates the axioms and rejects them; it is that the system proceeds as if the axioms were irrelevant. At that point, the epistemic status of the numbers is settled by convention, not by representation.

The Rasch-related items collapse uniformly to the floor (p = 0.05; −2.50), and this result is both expected and consequential. The proposition that transforming subjective responses to interval measurement is only possible with Rasch rules sits at the absolute minimum, as do the propositions that the Rasch logit ratio scale is the only basis for latent trait measurement, that the outcome of interest for latent traits is possession, and that Rasch rules are identical to representational measurement axioms. These results indicate that the EQ-5D knowledge base does not recognize the measurement problem posed by latent constructs and does not recognize Rasch measurement as the necessary solution when the attribute of interest is not directly observable.

In the context of EQ-5D, this matters because the instrument's descriptive responses are ordinal categories. Ordinal categories can represent ordering but not magnitude. The shift from three levels to five levels is often presented as improved sensitivity, reduced ceiling effects, and better discrimination. Yet increasing the number of ordinal categories does not change the fact that the categories remain ordinal. Without a transformation model that satisfies the axioms of measurement, the outputs cannot become interval or ratio quantities. The diagnostic's floor effect indicates that the knowledge base does not treat this as a foundational problem. Instead, it treats descriptive refinement as progress in measurement. This is a category error presented as methodological sophistication.

The item concerning possession of latent traits deserves special emphasis because it captures the ontological distinction the EQ-5D system systematically obscures. Measurement concerns attributes possessed by individuals. Valuation concerns preferences expressed by observers. The EQ-5D scoring architecture does not claim to measure a property intrinsic to the individual in the representational sense. It classifies health states descriptively and then assigns values to those states based on population preferences. The values reflect desirability under stated trade-offs, not possession of an attribute with invariant units. When analysts interpret changes in EQ-5D utilities as changes in the magnitude of "health-related quality of life," they treat preference-derived valuations as if they were measures of an attribute. The diagnostic result indicates that the knowledge base does not enforce the distinction between valuation and measurement as a governing constraint. Indeed, it is precisely the erasure of this distinction that makes the EQ-5D system usable as a numeric engine within cost-utility analysis.

The remaining five items, all at the floor, directly interrogate the epistemic core of preference-based scoring. The proposition that preference-based utilities create interval measures collapses completely. This is the central claim that would be required if the EQ-5D outputs are to support arithmetic operations beyond ordinal comparisons. Yet the knowledge base does not articulate an admissible mapping that preserves empirical structure in the manner required for interval measurement. What is typically offered instead is a practical rationale: preferences are elicited using formal techniques, responses are modeled, and an algorithm produces a number. But the existence of a number is not evidence of measurement. It is evidence of computation.

The proposition that ratio measures can have negative values collapses at the floor, and the EQ-5D system is a canonical illustration of why this matters; national value sets routinely allow negative values for some health states, with "states worse than dead" assigned values below zero. The instrument knowledge base then proceeds to treat these values as if they were compatible with ratio arithmetic in QALY construction and cost-effectiveness modeling. Under representational measurement theory, this is incoherent. A ratio scale requires a true zero and does not permit negative values. Negative values may be meaningful for preferences, utilities as expressions of desirability, or scores anchored to a reference state. They are not meaningful as magnitudes of an attribute with a true null. The floor effect here indicates that the EQ-5D environment has normalized the contradiction. It does not treat negative values as disqualifying for ratio arithmetic; it treats them as a routine feature of the system.

The proposition that multiattribute health state classifications are unidimensional collapses at the floor. This is decisive. It means the epistemic system does not recognize that combining heterogeneous domains into a single index is not measurement but aggregation. The EQ-5D's descriptive domains are not commensurate in the sense required for additive structure. Preference weighting imposes commensurability by fiat, not by demonstrating that the underlying empirical relations support a single quantitative continuum. The diagnostic outcome indicates that this imposition is not treated as an epistemic claim requiring justification. It is treated as a technical step required to produce a single number. In effect, the corpus treats the instrument's multiattribute structure as a feature, not a measurement disqualification.

The last two items, also at the floor, address the mechanism through which the category error is completed: the transformation of ordinal description into quantitative magnitude by preference weighting, and the assumption that preference algorithm scoring produces measurement-valid numerical quantities. Both propositions are false under the axioms of representational measurement, and the diagnostic indicates that the knowledge base fails to enforce this falsity as a constraint. Preference weighting can rank and assign relative desirability; it cannot produce invariant units of an attribute. Algorithmic scoring can compute a consistent output; it cannot establish that the output preserves empirical structure in the sense required for measurement. Yet within the EQ-5D system, the act of scoring becomes a proxy for the act of measuring. Numerical production is mistaken for quantitative representation.

Consider the internal coherence of the profile. The few items that are not at the absolute floor are still firmly negative and are exactly those items most likely to appear in the corpus as background language without functioning as constraints: true zero, unidimensionality, multiplication, and the sequencing of measurement and arithmetic. The pattern is therefore coherent: superficial recognition without disciplinary constraint. Once the diagnostic moves from definitional language to operational requirements—admissible arithmetic, Rasch transformation, and the status of preference scoring—the epistemic system collapses to complete non-possession.

This profile also clarifies why long-running disputes over valuation methods and mapping between versions do not constitute progress in measurement. The UK dispute between the EQ-5D-3L and EQ-5D-5L value sets, and the subsequent emphasis on mapping, are disputes internal to valuation practice. They concern how to generate preference-based numbers, which modeling strategy is preferred, and which set should be used in reference-case submissions. None of these disputes addresses the prior question of whether the output has the properties required for quantitative representation. The diagnostic profile explains why this question is rarely asked: it is not part of the epistemic grammar of the system. The system's internal debates occur entirely within the assumption that preference-based outputs are numerical quantities. Mapping then becomes a managerial device for compatibility, not a measurement argument.

The results have direct implications for what EQ-5D outputs can legitimately support. The instrument can provide structured descriptions of health-state profiles. It can support preference research about societal trade-offs. It can serve as a classificatory standard for summarizing domain responses. What it cannot support—absent a demonstration of unidimensionality, invariance, true zero, and admissible arithmetic—is the routine treatment of its index values as measures of magnitude, particularly when those values are averaged, differenced, multiplied by time, and aggregated. The existence of a value set does not resolve this; it is the mechanism through which valuation is institutionalized as pseudo-measurement.

A central feature of the EQ-5D epistemic system is that responsibility for foundational justification is systematically displaced. Developers can point to widespread adoption and methodological sophistication of valuation studies. Users can point to guidelines and reference-case requirements. Agencies can point to precedent and comparability. Educators can point to standard curricula. Each component defers the measurement question elsewhere, producing a closed loop in which numerical authority circulates without ever encountering admissibility constraints. The diagnostic does not merely criticize a set of practices; it exposes the epistemic architecture that sustains them.

The final point is structural invariance. The EQ-5D reduced-item profile is not an outlier. It reproduces the same pattern observed in the HUI and AQoL instrument-level interrogations. Despite differences in descriptive domains, response structures, and valuation protocols, the epistemic structure remains invariant: multiattribute ordinal description is converted to a single preference-weighted number; the number is treated as if it were a measure; arithmetic then proceeds as if admissible. The invariance across instruments is not coincidence; it is the predictable signature of a shared belief system. Instruments have been designed and institutionalized to feed a reference-case arithmetic architecture, not to satisfy the axioms that make arithmetic meaningful.

In that context, the significance of Table 1 is not that the EQ-5D system is poorly executed or that its users are inattentive to technical detail. The significance is that the system's governing knowledge base does not possess the conceptual conditions required for measurement. Where those conditions are absent, quantitative claims are not merely uncertain; they are undefined. Numerical storytelling prevails not because individuals choose to deceive, but because the epistemic system confers numerical legitimacy through repeated use rather than through representational demonstration.

The reduced canonical diagnostic therefore functions as a boundary test. It asks whether the instrument-level outputs satisfy the conditions required before any further arithmetic is attempted. The EQ-5D system fails that test comprehensively. The implication is unavoidable: disputes over which valuation method is preferable, which version is more sensitive, or how best to map between value sets are all disputes downstream of a prior and unresolved problem. Until measurement axioms are restored as admissibility conditions, technical refinements will continue to elaborate the machinery of pseudo-measurement while leaving the foundational category error intact.

If HTA is to claim scientific legitimacy, the epistemic order must be reversed. Measurement must precede arithmetic. Instruments must be evaluated not by how convenient they are for modeling but by whether they produce quantities under representational axioms. The reduced canonical profile for EQ-5D demonstrates that the UK reference-case ecosystem has not taken that step. Instead, it has institutionalized valuation as if it were measurement and then constructed a policy apparatus around the resulting numbers. The diagnostic makes visible what routine practice conceals: the numbers are treated as quantities because the system requires them to be quantities, not because they have been shown to be measures.

V CONCLUSIONS

The analysis presented in this working paper demonstrates that the EQ-5D-3L and EQ-5D-5L instruments do not fail measurement criteria in a partial, ambiguous, or transitional manner. They fail them structurally. The reduced canonical diagnostic reveals a stable and internally coherent epistemic profile in which the axioms required for quantitative measurement do not function as governing constraints within the knowledge base that authorizes the numerical use of EQ-5D outputs.

This failure is not attributable to implementation error, analytical misunderstanding, or insufficient technical sophistication. It reflects a deeper condition of non-possession. The principles that determine when numbers can meaningfully represent empirical attributes unidimensionality, invariant units, admissible scale type, and the logical precedence of measurement over arithmetic, do not operate as admissibility conditions within the EQ-5D epistemic environment. Where axioms are absent, they cannot constrain practice. Numerical operations may proceed, but they do so without representational authorization.

The transition from the EQ-5D-3L to the EQ-5D-5L does not alter this condition. Expansion in response levels refines ordinal description, but refinement within an ordinal framework does not establish measurement. Increased granularity does not generate invariant units, does not resolve multiattribute structure, and does not create a true zero. The epistemic architecture remains unchanged. What appears as methodological progress is, in fact, elaboration within a belief system that presupposes quantification rather than establishing it.

The reduced diagnostic makes clear that the EQ-5D system does not attempt to measure an attribute possessed by individuals in the representational sense. Instead, it assigns numerical values to multiattribute health-state descriptions using population-level preference elicitation. These values express desirability under hypothetical trade-offs. They do not express magnitude of an underlying attribute. When such valuations are subsequently treated as quantitative measures averaged, differenced, or multiplied by time the distinction between valuation and measurement collapses.

This collapse is not accidental. It is sustained by an epistemic system in which preference weighting is treated as a substitute for measurement rather than as a fundamentally different operation. Numerical form is mistaken for quantitative meaning. Algorithmic scoring is treated as conferring metric status. Over time, repeated use, institutional endorsement, and educational transmission normalize this substitution until the foundational question of measurement ceases to be asked at all.

Importantly, the diagnostic does not identify inconsistency or internal contradiction within the EQ-5D knowledge base. On the contrary, the endorsement profile is strikingly coherent. All propositions that would impose constraints on numerical interpretation are absent. All assumptions required to treat EQ-5D outputs as quantities remain unsupported. This coherence is itself diagnostic. It indicates not confusion, but epistemic closure. Where measurement axioms are not part of disciplinary grammar, they cannot be selectively applied.

The implications for inference are fundamental. Claims concerning magnitude of change, comparative therapeutic impact, or quantitative difference presuppose that numerical variation corresponds to variation in an underlying attribute. In the absence of representational measurement, such correspondence cannot be established. Apparent numerical change may therefore reflect the mechanics of scoring, weighting, and valuation rather than change in health itself. Precision becomes a property of computation, not of measurement.

This condition cannot be remedied through recalibration, alternative valuation techniques, or further refinement of descriptive systems. Measurement cannot be retrofitted. It must be established at the point of construction. As long as the EQ-5D remains a multiattribute classificatory system whose outputs are generated through preference algorithms, the structural requirements for quantitative measurement cannot be satisfied.

The findings therefore locate the failure not in the instrument's execution, but in its epistemic role. The EQ-5D functions exactly as designed: as a standardized vehicle for preference valuation. The error arises only when outputs from such a system are elevated to quantitative status and used as if they were measures of magnitude. Valuation is not measurement, and numerical convenience cannot substitute for representational validity.

This conclusion aligns directly with findings from instrument-level interrogations of the HUI and the AQoL families. Despite differences in descriptive domains, valuation protocols, and national origins, the epistemic structure is invariant. Each system occupies the same conceptual position: a preference-based valuation framework embedded within a knowledge base that treats valuation as measurement in order to sustain downstream arithmetic.

Recognizing this does not require abandonment of descriptive classification or preference research. It requires epistemic realignment. Instruments such as the EQ-5D may retain value as classificatory or comparative tools, but only if their numerical outputs are demoted from quantitative status. Claims about magnitude, change, and comparative effect must be restricted to what the instrument can legitimately support.

Until measurement axioms are restored as admissibility conditions for numerical claims, health technology assessment will continue to operate within a closed epistemic loop in which numbers circulate without representation. The present analysis renders that condition explicit. In doing so, it establishes not merely a critique of the EQ-5D, but a foundation for re-thinking what it means to measure health at all.

This conclusion does not require erasing the instrument from use entirely. Multiattribute descriptive systems such as the EQ-5D may retain limited classificatory or descriptive value, for example in summarizing health state profiles or facilitating structured reporting. However, this requires a clear epistemic demotion. Such instruments must no longer be treated as quantitative measures of magnitude or change.

The more defensible path forward lies in restoring measurement as a precondition for arithmetic. Outcome instruments must either be constructed explicitly to satisfy the axioms of representational measurement or be restricted to inferential claims consistent with their descriptive nature. Without this realignment, therapy impact claims derived from EQ-5D scores will continue to present numerical form without quantitative substance.

Recognizing the foundational role of representational measurement does not entail abandoning quantitative evaluation of health outcomes. It requires reinstating logical discipline in the relationship between numbers and the attributes they purport to represent. Only through such discipline can health technology assessment move from numerical convention toward empirically meaningful knowledge.

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] Buchholz I, Mathieu F, Janssen M et al. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D." *PharmacoEconomics* 36, no. 6 (2018): 645–661

[2] Shaw C, Longworth L, Bennett B,. A Review of the Use of EQ-5D for Clinical Outcome Assessment in Health Technology Assessment, Regulatory Claims, and Published Literature." *Patient*, 2024; 17(30): 251

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

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