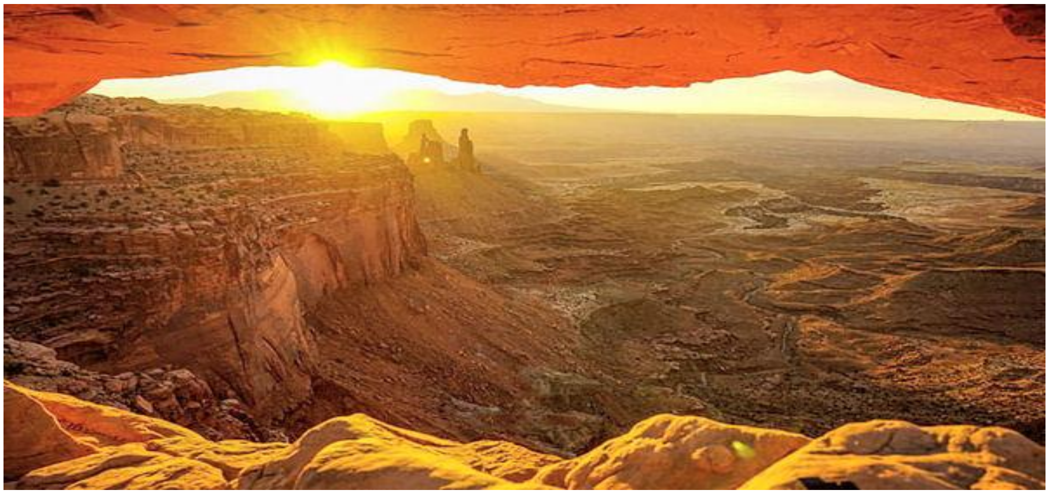
MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION



REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

CANADA: DECONSTRUCTING THE EPISTEMIC KNOWLEDGE BASE OF THE HUI INSTRUMENTS

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

LOGIT WORKING PAPER No 60 FEBRUARY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This working paper deconstructs the epistemic knowledge base that authorizes use of the Health Utilities Index (HUI) as if its outputs were quantitative measures. In health technology assessment (HTA), instruments are often treated as neutral devices whose numerical outputs acquire meaning at the point of construction and validation. This paper rejects that assumption. Numerical legitimacy is socially conferred: numbers become treated as quantities only when communities repeatedly subject them to arithmetic and accept the outputs as meaningful. Over time, routine practice, institutional endorsement, methodological repetition, and education transform assumption into norm, rendering measurement questions epistemically invisible.

The HUI provides a revealing case. Developed in Canada as a multiattribute health-state classification system valued through preference elicitation, the HUI family is routinely deployed across clinical research, economic evaluation, and policy analysis. Its outputs are treated as "utilities," summarized through means and differences, multiplied by time to generate QALYs, and embedded in models and software templates. This paper argues that such usage does not demonstrate measurement; it substitutes numerical manipulation for measurement justification.

The analysis treats the HUI not as an isolated technical artifact but as an epistemic object embedded within a distributed user epistemic system. That system includes developers, applied researchers, reviewers and journals, HTA agencies and guideline authors, educators, and analytic infrastructure that automates scoring and embeds assumptions in code. The central question is whether this knowledge base recognizes representational measurement axioms as admissibility conditions for arithmetic.

To address this, the paper interrogates the HUI corpus using a reduced canonical diagnostic grounded in representational measurement theory and Rasch requirements for latent traits. Endorsement probabilities are assigned as epistemic classifications of reinforcement and converted to normalized logits for profile interpretation. The results yield a stable and internally coherent pattern of uniformly negative logits, indicating structural non-possession of measurement axioms within the HUI knowledge base. The implications are direct: HUI outputs function as preference-based valuations rather than measures, and downstream arithmetic uses—including QALY construction, inherit this non-measurement. The paper concludes that technical refinement cannot repair a foundational category error: valuation does not become measurement through repetition.

I INTRODUCTION: THE HUI USER EPISTEMIC SYSTEM

In health technology assessment, instruments are commonly treated as neutral devices whose numerical outputs are assumed to possess inherent meaning once they have been developed and validated. Under this view, the epistemic work of measurement is believed to occur at the point of construction. Thereafter, numbers are assumed to be portable, stable, and suitable for arithmetic wherever they are applied. Analytical attention is therefore directed toward how numbers are used rather than whether they measure anything at all. This assumption is deeply misleading. Instruments do not acquire numerical legitimacy through construction alone. They acquire it through use.

Numbers come to be treated as quantities only when a community repeatedly subjects them to arithmetic operations and accepts the resulting outputs as meaningful. That acceptance is rarely explicit. It emerges gradually through routine practice, institutional endorsement, methodological repetition, and educational transmission. Over time, what begins as assumption becomes belief, and belief becomes norm. Once embedded in this way, the question of whether an instrument actually measures an empirical attribute ceases to appear as a scientific problem. It becomes epistemically invisible.

In other words, the question of the scales of measurement defining allowable arithmetic operations developed and published by Stevens in 1946, the Rasch rules for transforming observations capturing latent traits to a logit ratio scale published in 1960 and the formalization of the axioms of representational measurement in 1971 by Krantz et all fail to be in the possession of those who developed and those who apply the Health Utilities Index (HUI) [1 2 3 4]. The proposition, or more formally the rule, that measurement should precede arithmetic is an entirely foreign concept although accepted some 80 years ago.

The HUI provides a particularly revealing illustration of this inversion process [5 6]. Originally developed in Canada as a system for classifying health states and assigning preference-based values to those states, the HUI family of instruments has become embedded across clinical research, health economics, and policy evaluation. Its numerical outputs are routinely interpreted as utilities, incorporated into economic models, multiplied by time to generate quality-adjusted life years, and aggregated across individuals and populations. These operations are not exploratory or provisional. They are routine. Through repetition, numerical manipulation has come to substitute for measurement justification.

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 may 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 prevails.

The object of analysis is therefore not the HUI instrument in isolation, nor the intentions of its developers, nor the technical details of its valuation procedures. The object of analysis is the epistemic system within which HUI outputs function as if they were quantitative measures. This paper refers to that system as the user epistemic system. It is this system, not the instrument itself that confers numerical authority.

The user epistemic system consists of the collective practices through which HUI values are interpreted as meaningful numbers. It includes researchers who analyze HUI data, reviewers who evaluate manuscripts, journals that publish results, health technology assessment agencies that accept HUI 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 authoritative declaration that HUI 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 analytically routine. Collectively, they construct a powerful presumption: if numbers are used as quantities, they must be quantities. Use becomes evidence; repetition becomes validation.

Within such a system, epistemic responsibility becomes diffuse. Developers may point to widespread application. Users may point to methodological convention. Agencies may point to precedent. Educators may point to established 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.

The distinction between ignorance and non-possession is therefore central. The problem is not that users of the HUI misunderstand 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 evolution of the HUI family illustrates this dynamic clearly. The development of successive versions, including refinements in attribute definitions and valuation algorithms, is commonly interpreted as methodological progress. Yet such refinements operate entirely within an already accepted numerical framework. They presuppose that the outputs are measures rather than establishing that they are. Technical elaboration thus occurs without epistemic transformation.

This paper therefore does not ask whether the HUI is useful, practical, or historically influential. It asks a more fundamental question: does the knowledge base that authorizes its numerical use contain the axioms required for measurement? By treating the HUI 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 HUI 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 for 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 HUI 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 HUI system. This cannot be limited to the original development papers produced by the HUI research group, nor can it be confined to technical documentation describing scoring algorithms or valuation procedures. Once introduced into applied research and policy contexts, the HUI became embedded within a much 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 numerical outputs as quantitative measures. The HUI knowledge base must therefore be understood as an epistemic corpus.

This corpus includes the foundational publications describing HUI2 and HUI3, which established the instrument's conceptual architecture. The HUI was designed as a multiattribute health state classification system, incorporating domains such as vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Each attribute is expressed through ordered categorical levels, which together define a large space of possible health states. From its inception, the instrument was framed not as a direct measure of health, but as a system for describing health states to be assigned values through preference elicitation. This framing is central. It positions valuation, rather than measurement, as the conceptual foundation of the instrument.

The authority of the HUI does not persist because these development papers are repeatedly examined or theoretically defended. It persists because subsequent users treat the resulting numerical outputs as if they were quantities. The dominant component of the HUI knowledge base therefore lies in applied research. Across clinical trials, population surveys, and health economic evaluations, HUI scores are reported as outcomes. Means are calculated, changes interpreted, and differences compared across interventions and groups. These operations are typically presented without discussion of scale type, unidimensionality, or invariance. Yet the absence of such discussion is itself epistemically consequential. It signals that justification is unnecessary.

Health technology assessment agencies constitute a particularly influential layer of the HUI knowledge base. In Canada and in several international contexts, HUI utilities have been accepted as legitimate inputs for economic evaluation. This acceptance confers institutional authority. It does not arise from demonstration that HUI values satisfy the axioms of representational measurement, but from their compatibility with established evaluative frameworks. Once incorporated into agency guidance or accepted practice, the instrument's numerical status becomes administratively secured rather than theoretically established.

Methodological documents further reinforce this authority. Economic evaluation guidelines, submission templates, and analytic manuals frequently treat HUI utilities as interchangeable numerical entities. They are abstracted from their descriptive origins and integrated into standardized workflows. At this stage, the instrument ceases to function as an object of epistemic inquiry and becomes instead a component of analytic infrastructure. The question of what kind of numbers HUI produces is displaced by the assumption that numbers are required.

Education plays a decisive role in reproducing this knowledge base. In graduate programs and professional training, students encounter HUI values as conventional inputs. They learn how to apply utilities, how to calculate quality-adjusted life years, and how to interpret incremental ratios. Rarely are they taught to interrogate whether the utilities themselves possess the properties required for arithmetic. By the time analysts enter professional practice, numerical legitimacy has already been internalized. The instrument is encountered not as a theoretical proposition, but as a methodological given.

The epistemic reach of the HUI knowledge base extends further through analytic infrastructure. Software packages, statistical routines, and economic models embed HUI scoring functions and value sets. Once encoded, the conceptual assumptions underlying the instrument become invisible. Users interact with numerical outputs without encountering the premises that authorize their use. In this way, epistemic commitment is no longer expressed through argument or citation, but through automation.

The existence of multiple HUI versions illustrates how the knowledge base absorbs modification without altering its foundations. Differences between HUI2 and HUI3 are commonly interpreted as technical refinement, improved descriptive coverage, or enhanced sensitivity. Yet these modifications do not alter the underlying epistemic architecture. The instrument remains multiattribute. Valuation remains preference-based. Utilities continue to permit negative values. Arithmetic compatibility with cost-utility analysis remains assumed rather than demonstrated. The proliferation of versions therefore represents elaboration within a stable belief system rather than epistemic transformation.

Crucially, the HUI knowledge base is not unified by explicit theoretical agreement. There is no authoritative statement asserting that HUI utilities satisfy representational measurement axioms. Instead, unity arises 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 HUI system. Developers can point to extensive use. Users can point to agency acceptance. 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 HUI knowledge base in this way is essential for the analysis that follows. The purpose of interrogation is not to assess individual publications or authorial 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.

The following section therefore turns to interrogation of the HUI knowledge base using a reduced canonical diagnostic grounded in representational measurement theory and Rasch principles. This interrogation does not ask whether the HUI is widely used or technically sophisticated. It asks whether the system that authorizes its use possesses the conditions required for measurement itself.

III INTERROGATING THE HUI KNOWLEDGE BASE

If the numerical authority of the Health Utilities Index derives from a distributed epistemic environment rather than from instrument construction alone, then evaluating the HUI requires a method capable of interrogating that environment. The object of analysis is not whether individual studies correctly apply the instrument, nor whether its developers intended to create a measure. The relevant question is whether the knowledge base that authorizes the numerical use of HUI outputs 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 evaluate methodological competence or good faith. Instead, interrogation examines what the knowledge base articulates, normalizes, or excludes. It identifies which principles function as operative constraints and which are absent from disciplinary practice.

The theoretical foundation for this approach 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 structures. Scale type determines permissible arithmetic operations. Ordinal data support ordering only. Interval scales permit addition and subtraction but lack a meaningful zero. Ratio scales alone permit multiplication, division, and comparison of magnitudes. These distinctions are not methodological preferences but logical preconditions for quantification. Where they are not satisfied, arithmetic operations are undefined regardless of how routinely they are applied in practice.

The requirements become more stringent when the attribute of interest is latent. Constructs such as health status or functioning are not directly observable and must be inferred from response patterns. In such cases, ordinal observations cannot be treated as quantities unless transformed through an explicit measurement model capable of producing invariant units. Rasch measurement provides such a model. It is the only framework that simultaneously satisfies the axioms of representational measurement and yields a linear logit ratio scale suitable for quantitative comparison. Without such transformation, numerical scores remain ordinal regardless of analytic sophistication.

The interrogation framework employed here therefore draws upon a canonical set of propositions derived from representational measurement theory and Rasch measurement principles. Each proposition expresses either a necessary condition for measurement or a known impossibility when those conditions are violated. Examples include the requirement of unidimensionality, the incompatibility of negative values with ratio scales, and the necessity of invariant transformation for latent traits. These propositions do not represent theoretical preferences; they articulate the logical structure of measurement itself.

For purposes of instrument-level evaluation, a reduced canonical diagnostic is employed rather than the full twenty-four item framework used in system-level assessments. This reduction is deliberate and principled. Instruments do not authorize policy arithmetic, aggregation across populations, or cost-effectiveness modeling. Their epistemic responsibility lies upstream. They generate numerical outputs that may later be subjected to arithmetic by other actors. Including system-level propositions at this stage would therefore introduce category error, attributing downstream analytic claims to upstream representational structures.

The reduced canonical diagnostic is accordingly confined to propositions that test the ontological and representational status of the instrument's outputs prior to any modeling or policy application. It interrogates whether the knowledge base recognizes the conditions under which numerical outputs could, in principle, function as measures. This establishes what may be termed the pre-arithmetic boundary. 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 if, and only if, they articulate necessary conditions for measurement under representational measurement theory. These include propositions concerning unidimensionality, scale type, the existence of a true zero, invariance, and the permissibility of arithmetic operations. Conversely, statements are excluded if they presuppose that measurement has already occurred. Items concerning aggregation, time-based multiplication, or composite outcome construction are downstream claims whose validity depends entirely on whether quantities exist in the first place.

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

The interrogation does not assess whether these propositions are explicitly endorsed or rejected in published texts. It examines whether they function as operative constraints within the knowledge base. A principle may never be denied 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 patterns should therefore not be interpreted as error or confusion. They represent the expected signature of structural non-possession.

The reduced canonical diagnostic thus functions as a structural probe. It reveals whether the epistemic environment contains the conceptual architecture required for measurement to occur. Where that architecture is absent, numerical practice may remain widespread, institutionally endorsed, and methodologically elaborate, yet still lack representational validity.

IV RESULTS

The results presented in Table 1 is not surprising, but its consistency is nonetheless instructive. The HUI has long been treated as a technically sophisticated instrument, widely adopted across clinical research, population health studies, and health economic evaluation. Its numerical outputs are routinely reported as utilities, summarized through means, compared across groups, and interpreted as magnitudes of health-related quality of life. Yet the diagnostic demonstrates that these practices do not arise from an underlying recognition of measurement axioms. They arise from convention. Numerical treatment has become normalized in the absence of explicit measurement justification.

TABLE 1: REDUCED ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS HUI INSTRUMENTS CANADA

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	-2.20
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 TRANSFORMED INTO QUANTITATIVE MAGNITUDE	0	0.05	-2.50

THROUGH PREFERENCE WEIGHTING			
PREFERENCE ALGORITHM SCORING PRODUCES MEASUREMENT VALID NUMERICAL QUANTITIES	0	0.05	-2.50

The first group of results concerns the most elementary distinction in measurement theory: the difference between interval and ratio scales. The proposition that interval measures lack a true zero registers an endorsement probability of 0.20, corresponding to a normalized logit of −1.40. This indicates weak and inconsistent reinforcement. While the distinction is not wholly absent from the knowledge base, it does not function as a governing constraint on analytic practice. In the HUI environment, numerical outputs are routinely treated as if zero represented the absence of health, even though no empirical structure is offered to justify such an interpretation.

This ambiguity is consequential. A true zero is not a semantic convenience but a defining property of ratio measurement. It anchors arithmetic interpretation by representing the absence of the attribute being measured. Without a true zero, multiplication and division are undefined. The low endorsement observed here indicates that this foundational requirement does not structure how HUI values are interpreted or used. Zero functions operationally, not representationally. It is accepted because it appears numerically convenient, not because it has been shown to correspond to an empirical null state.

The diagnostic collapses further when attention turns to unidimensionality. The proposition that measures must be unidimensional receives an endorsement probability of 0.10, yielding a normalized logit of −2.20. This result reflects near-total absence of reinforcement. Unidimensionality does not function as a constraint within the HUI epistemic system. This is not an incidental omission. The HUI is explicitly multiattribute by construction, combining domains such as vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain through preference-weighted aggregation. These domains are qualitatively heterogeneous and lack a common empirical unit.

From the perspective of representational measurement theory, this heterogeneity is decisive. Measurement requires that observed variations correspond to variation along a single attribute. Where multiple attributes are combined, the result is not a measure of anything in particular, but a composite index whose numerical value reflects weighting choices rather than empirical magnitude. The diagnostic indicates that this distinction is not recognized as problematic within the HUI knowledge base. Multiattribute aggregation is treated as compatible with measurement, despite violating the most basic requirement for quantitative representation.

Closely related to unidimensionality is the proposition that measurement must precede arithmetic. This item also registers at 0.10 (−2.20). The implication is clear: arithmetic operations are undertaken without prior establishment of measurement properties. In the HUI system, numerical manipulation does not follow measurement; it substitutes for it. Values are added, averaged, and compared not because their scale properties permit such operations, but because analytic frameworks require numbers to function.

This inversion represents a profound epistemic shift. Under representational measurement theory, arithmetic is conditional. It is permitted only when the empirical structure of the attribute supports the corresponding numerical operation. In the HUI environment, this conditionality is absent. Arithmetic becomes a methodological default rather than a logically constrained act. Numbers are treated as quantities by virtue of their numerical form alone.

The same pattern appears in the endorsement of the proposition that multiplication requires a ratio measure. With an endorsement probability of 0.10, this principle does not function as a constraint within the knowledge base. The implication is not merely technical. It indicates that scale type is not recognized as determining what mathematical operations are permissible. Whether a quantity can be meaningfully multiplied is treated as a modeling decision rather than a measurement question.

Taken together, these four items reveal a consistent epistemic orientation. The HUI knowledge base does not operate under the logic that measurement governs arithmetic. Instead, arithmetic governs measurement. Numerical operations define what counts as a quantity, rather than the reverse. This inversion is not explicitly defended; it is simply assumed. Over time, repeated practice has rendered the assumption invisible.

The diagnostic reaches its most decisive results when it turns to the measurement of latent constructs. All Rasch-related propositions collapse to the absolute floor of the scale, with endorsement probabilities of 0.05 and normalized logits of −2.50. The knowledge base does not recognize Rasch transformation as necessary for converting ordinal observations into interval measures. It does not recognize the Rasch logit ratio scale as the only defensible basis for latent trait measurement. Nor does it recognize the equivalence between Rasch axioms and representational measurement theory.

These findings are structurally important. Latent constructs cannot be observed directly. Their measurement requires a model that establishes invariant relations between persons and items. Without such a model, responses remain ordinal, regardless of how they are coded or weighted. The HUI framework does not include such a transformation. Instead, it relies on preference elicitation to assign numerical values to health states. Preferences, however, express order and desirability, not magnitude. They cannot, by themselves, generate interval or ratio measures.

The collapse of Rasch-related items therefore reflects not a methodological oversight but a categorical incompatibility. Rasch measurement requires unidimensionality and invariance. The HUI framework explicitly violates both. Its multiattribute structure precludes unidimensional measurement, and its reliance on external valuation undermines invariance at the individual level. The diagnostic does not reveal an absence that could be remedied; it reveals a structural impossibility.

This incompatibility becomes even clearer in the endorsement of the proposition that the outcome of interest for latent traits is possession of that trait. With an endorsement probability of 0.05, the knowledge base does not conceptualize health-related quality of life as an attribute possessed by individuals. Instead, individuals are located within a classificatory space of health states that are valued externally by societal preferences. The numerical output reflects how a state is valued, not how much of an attribute an individual possesses.

This distinction is fundamental. Measurement concerns attributes of entities. Valuation concerns preferences of observers. The HUI operates entirely within the latter domain. Its numerical outputs represent collective judgments about the desirability of hypothetical states, not magnitudes of a property inherent in the individual. Treating such values as measures of individual health therefore entails a category error that the epistemic system does not acknowledge.

The diagnostic further demonstrates that preference-based scoring does not function as a recognized pathway to measurement. The proposition that preference-based utilities create interval measures collapses completely, with an endorsement probability of 0.05. Despite decades of use, the knowledge base does not articulate or defend the claim that preference elicitation yields metric quantities. Instead, interval properties are assumed by convention. The existence of a numerical algorithm is taken as sufficient justification.

This assumption is reinforced by the equally strong rejection of the proposition that ratio measures can have negative values. Negative utilities are an explicit feature of the HUI framework, allowing health states to be valued as worse than dead. Yet under representational measurement theory, negative values are incompatible with ratio scales, which require a true zero representing absence of the attribute. The diagnostic indicates that this contradiction is not treated as problematic. It is absorbed into practice without theoretical reconciliation. This ratio scale requirement was made clear by Stevens in 1946 in his seminal paper on scales of measurement.

The presence of negative values is particularly revealing because it exposes the conflation of valuation and measurement. Preferences can extend below a reference point, expressing aversion or undesirability. Measurements cannot. A negative mass, negative length, or negative temperature on an absolute scale is meaningless. By permitting negative utilities while simultaneously treating those utilities as quantitative magnitudes, the HUI framework collapses the distinction between desirability and quantity. The epistemic system accepts this collapse without resistance.

The proposition that multiattribute health-state classifications are unidimensional likewise collapses to the floor. This result is unsurprising, yet its implications are profound. The HUI classification system explicitly comprises multiple dimensions that do not share a common empirical unit. Aggregation across these domains is achieved through preference weighting, not through demonstration of additive structure. The resulting index is therefore not a measure of any single attribute. It is a numerical summary of heterogeneous judgments.

From a representational measurement standpoint, such aggregation is illegitimate. Additivity requires that concatenation of the attribute be empirically meaningful. No such concatenation exists across vision, hearing, cognition, and pain. The diagnostic indicates that this absence does not constrain practice. Aggregation is treated as a technical step rather than a theoretical claim.

The final two items address the transformation from ordinal description to quantitative magnitude. Both collapse entirely. The knowledge base does not recognize that ordinal health-state descriptions cannot be transformed into quantitative magnitude through preference weighting alone. Nor does it recognize that algorithmic scoring confers numerical form without conferring measurement validity. These results strike at the heart of the instrument's epistemic status.

Preference algorithms can assign numbers. They cannot establish that those numbers preserve empirical relations in the attribute of interest. Without such preservation, numerical differences lack substantive meaning. The diagnostic reveals that this distinction is not operative within the HUI epistemic system. Numbers are treated as quantities because they behave numerically, not because they represent quantities.

When viewed as a whole, the endorsement profile exhibits complete internal coherence. There are no contradictions between items, no mixed signals, and no evidence of partial conceptual transition. All propositions that would impose constraints on numerical use are rejected or absent. All assumptions required to treat HUI outputs as measures are unsupported. The pattern is stable and systematic.

This uniformity is diagnostically important. It demonstrates that the failure observed is not the result of isolated misconceptions or disciplinary disagreement. It reflects a fully consolidated epistemic system in which measurement theory does not function as a governing authority. The absence of axioms produces consistent outcomes. Where constraints are not recognized, none can be selectively applied.

Importantly, these results should not be interpreted as an indictment of individual researchers, clinicians, or analysts. The diagnostic does not assess beliefs or competence. It assesses possession. The principles tested do not operate as admissibility conditions within the knowledge base. Where axioms are not part of disciplinary grammar, even highly trained analysts cannot invoke them.

The implications of this finding extend beyond the HUI itself. The same endorsement profile has been observed for other preference-based multiattribute instruments, including the EQ-5D and AQoL families. Despite differences in descriptive systems, valuation protocols, and national origins, the epistemic structure is invariant. Each instrument is embedded within a knowledge base that treats valuation as measurement and numerical convenience as quantitative legitimacy.

This invariance is not accidental. It reflects the historical trajectory of health utility theory, which prioritized comparability and model compatibility over measurement foundations. Instruments were designed to feed analytic frameworks rather than to establish quantities. Over time, those frameworks became institutionalized, and their numerical requirements came to define what counted as measurement.

The diagnostic therefore reveals a deeper epistemic condition. The HUI does not fail because it is poorly designed. It fails because it was never designed to measure in the representational sense. Its purpose is classificatory and valuational, not metrical. The epistemic error arises when outputs from such a system are treated as magnitudes of health rather than expressions of preference.

Recognizing this distinction does not imply that the HUI lacks all value. As a descriptive system, it can facilitate structured reporting of health states. As a valuation framework, it can inform preference studies. What it cannot do is support quantitative inference about magnitude, change, or comparative effect. Those claims require measurement properties that the instrument does not and cannot possess.

The results of the reduced canonical diagnostic therefore support a clear conclusion. The HUI functions within an epistemic system that does not recognize the axioms required for measurement. Its numerical outputs are treated as quantities by convention rather than by demonstration. The resulting arithmetic practices rest on belief rather than representation.

This condition cannot be corrected through refinement, recalibration, or expansion of the instrument family. Adding domains, modifying weights, or increasing descriptive granularity does not address the absence of unidimensionality, invariance, and true zero. Measurement cannot be added after the fact. It must be established at the point of construction.

The significance of this analysis lies not in critiquing a single instrument, but in clarifying the conditions under which numerical authority has been constructed in health outcomes research. By exposing the epistemic foundations of that authority, the diagnostic makes visible what routine practice has rendered invisible. Numbers have come to stand in for quantities without satisfying the conditions that make such representation meaningful.

In this sense, the HUI is not an outlier. It is exemplary. It illustrates how numerical systems can acquire institutional legitimacy in the absence of measurement, sustained through repetition, endorsement, and educational transmission. The challenge for future work is not to improve such instruments, but to reconsider the epistemic standards by which quantitative claims are authorized.

Only by restoring measurement axioms as admissibility conditions can health outcomes research move beyond numerical storytelling toward genuinely quantitative science.

V CONCLUSIONS

The analysis presented in this working paper demonstrates that the HUI does not fail measurement criteria sporadically, partially, or ambiguously. It fails 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 sustains the instrument's use.

This failure is not attributable to deficiencies in implementation, analyst error, or misunderstanding of technical detail. It reflects the deeper condition of non-possession. The principles that determine when numbers can meaningfully represent empirical attributes are not operative within the epistemic environment in which the HUI functions. Where axioms are absent, they cannot constrain practice. Numerical operations may continue, but they do so without logical authorization.

The implications of this finding are fundamental. The HUI does not generate quantities that preserve empirically testable relational structure. Its outputs do not satisfy unidimensionality, do not possess invariant units, lack a true zero, and are derived through preference-based aggregation across heterogeneous attributes. These characteristics are not minor limitations. They preclude measurement in the representational sense altogether.

Importantly, this conclusion does not imply that the HUI is poorly designed or inadequately calibrated. The instrument performs exactly the function it was constructed to perform: the valuation of multidimensional health-state descriptions using population preferences. The epistemic error arises only when such valuations are subsequently treated as quantitative magnitudes of health. Valuation is not measurement, and numerical form does not confer quantitative meaning.

The findings therefore locate the failure not within the instrument alone, but within the epistemic system that authorizes its numerical interpretation. Through repeated use, institutional endorsement, and educational transmission, HUI outputs have come to be treated as if they were measures despite lacking the properties required for measurement. This transformation does not occur through explicit theoretical justification, but through normalization. Arithmetic replaces axioms; convention replaces representation.

This pattern is not unique to the HUI. The diagnostic profile observed here mirrors those found for other preference-based multiattribute instruments, including the EQ-5D and AQoL families. Despite differences in descriptive architecture, valuation protocol, and national origin, the epistemic structure remains invariant. Each instrument occupies the same conceptual position: a valuation system embedded within an epistemic environment that treats valuation as measurement.

For this reason, technical refinement cannot resolve the problem. Modifying weights, expanding response levels, or introducing new versions does not address the absence of unidimensional structure, invariant units, or lawful arithmetic permission. Measurement cannot be retrofitted. It must be established at the point of construction.

The implications for health technology assessment are unavoidable. When instruments that do not measure are treated as quantitative inputs, all downstream analyses inherit that non-measurement. No degree of modeling sophistication, statistical adjustment, or policy formalism can compensate for the absence of quantity at the source. Apparent precision is thereby transformed into numerical storytelling.

Recognizing this condition does not require abandoning structured description or preference research. It requires epistemic realignment. Instruments such as the HUI may retain descriptive or classificatory value, but they must be demoted from quantitative status. Claims about magnitude, change, or comparative effect must be restricted to what the instrument can legitimately support.

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

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

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

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

[5] Horsman J. Furlong F, Feeny D, Torrance G. The Health Utilities Index (HUI®): Concepts, Measurement Properties and Applications.*" Health and Quality of Life Outcome.* 2003; 1(54)

[6] Bolbocean, C, Anderson P, Bartmann P et al. *C*omparative Evaluation of the Health Utilities Index Mark 3 and Short Form-6D: Evidence from Individual Participant Data Meta-Analysis. *Quality of Life Research.* 2023; 32: 1703–1716.