

MAIMON RESEARCH LLC
**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**EPISTEMIC STRUCTURAL INVARIANCE IN THE
FALSE MEASUREMENT STATUS OF NATIONAL
MULTIATTRIBUTE INSTRUMENTS:**

**AUSTRALIA, CANADA, FINLAND AND THE UNITED
KINGDOM**

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

LOGIT WORKING PAPER No 77 JANUARY 2026

www.maimonresearch.com

Tucson AZ

EPISTEMIC STRUCTURAL INVARIANCE IN THE FALSE MEASUREMENT STATUS OF NATIONAL MULTIATTRIBUTE INSTRUMENTS: AUSTRALIA, CANADA, FINLAND AND THE UNITED KINGDOM

ABSTRACT

Over the past five decades, health technology assessment (HTA) has institutionalized the use of multiattribute preference-based instruments to quantify health-related quality of life and to support cost-utility analysis. Instruments such as the EQ-5D, the Health Utilities Index (HUI), the Assessment of Quality of Life (AQoL), and the Finnish 15D differ in descriptive structure, national origin, and valuation protocol, yet they are routinely treated as if they generate comparable quantitative outputs. Debate within HTA has therefore centered on technical refinements—sensitivity, responsiveness, ceiling effects, and national value sets—while presupposing that the numerical outputs already possess the properties required for arithmetic. The prior question, whether these instruments produce measures at all, is typically left unasked.

This paper addresses that omission by locating instrument legitimacy upstream of modeling and policy arithmetic and interrogating the epistemic systems that authorize instrument use. Building on the Logit Working Papers series, it applies canonical diagnostics derived from representational measurement theory and Rasch principles to determine whether the conceptual conditions necessary for numerical representation operate as admissibility constraints within the knowledge bases sustaining these instruments. The diagnostics are not psychometric performance tests; they do not evaluate empirical “fit,” reliability, or implementation competence. Instead, they probe reinforcement of foundational measurement propositions—unidimensionality, scale-type constraints, invariance of units, the rule that measurement must precede arithmetic, and the requirement that latent attributes be transformed to an invariant logit ratio scale through Rasch measurement if measurement is to occur at all.

A central finding motivates the present analysis: reduced canonical statement profiles converge across instruments developed independently in different countries and embedded within different institutional traditions. This convergence is interpreted as evidence of epistemic structural invariance—a shared belief structure that governs what counts as “measurement” in HTA irrespective of instrument design. The paper argues that this invariance reflects not mere misunderstanding but non-possession: measurement theory does not function as a governing authority within these systems. Where admissibility axioms are absent, numerical practice proceeds not by violating rules, but in a conceptual space where the rules do not exist.

Rasch invariance provides the keystone analogy. In Rasch measurement, invariance is a defining condition separating measurement from scoring: item parameters must be stable across populations and person measures stable across item sets, permitting interpretation of differences as differences in attribute possession rather than context dependence. Multiattribute instruments fail at the first step because measurement presupposes a single attribute; aggregating heterogeneous domains yields an index, not a quantity with invariant units. Preference weighting thus substitutes valuation for measurement, allowing arithmetic to proceed without representational grounding. The paper concludes that incremental reforms—tariff updates,

expanded response levels, refined valuation protocols, or more elaborate modeling—cannot introduce invariance after the fact. Measurement must be foundational, and where it is absent everywhere, non-measurement becomes the invariant structure of HTA practice.

INTRODUCTION

Over the past five decades, health technology assessment has come to rely heavily on multiattribute standardized instruments intended to quantify health-related quality of life. These instruments are widely treated as interchangeable components within economic evaluation frameworks, permitting comparison of therapies, aggregation of outcomes, and calculation of cost-effectiveness metrics. Although they differ in descriptive structure, national origin, and valuation protocol, instruments such as the EQ-5D, the Health Utilities Index (HUI), the Assessment of Quality of Life (AQoL), and the Finnish 15D are routinely assumed to generate comparable quantitative outputs.

This assumption of equivalence has rarely been interrogated at the level that matters most: whether these instruments produce measures at all. Most debates focus on sensitivity, responsiveness, ceiling effects, or national value sets, presupposing that the numerical outputs already possess legitimate quantitative properties. Measurement itself is treated as settled.

The Logit Working Papers series has challenged this presupposition by applying canonical diagnostics derived from representational measurement theory and Rasch principles to the epistemic systems that authorize instrument use. These diagnostics do not assess psychometric performance. They test whether the conceptual conditions required for numerical representation are present within the knowledge base that sustains each instrument.

A striking result has emerged. When reduced canonical statement sets are applied separately to national instruments developed in different countries, by different research groups, and under different policy traditions, the endorsement profiles converge. The pattern is not approximate. It is structurally invariant.

This paper examines that convergence. It argues that the repeated appearance of identical reduced-item profiles across countries constitutes evidence of epistemic structural invariance: a shared underlying belief system governing what counts as measurement in health technology assessment, independent of instrument design or national context.

STRUCTURAL INVARIANCE AS AN EPISTEMIC CONCEPT

In measurement science, invariance occupies a foundational role. A numerical scale qualifies as a measure only if it preserves meaning across contexts. In Rasch measurement, this requirement is formal and non-negotiable. Item difficulty parameters must remain stable across different populations, and person measures must remain stable across different item sets. If item calibrations shift when the sample changes, or if person measures depend on the particular questions asked, then the scale does not represent an underlying attribute. It represents context. Without invariance, comparison collapses. A scale that behaves differently across settings is not a measure but a scoring convention.

This requirement distinguishes Rasch measurement from all other psychometric traditions. Classical test theory and most item response models tolerate context dependence and sample-specific calibration. Rasch does not. Invariance is not an empirical aspiration but a logical condition. It is the mechanism by which numbers become representations of an attribute rather than summaries of responses. Only when invariance holds can differences be interpreted as differences in possession of the attribute being measured, rather than differences in instrument structure or population composition.

The importance of this requirement cannot be overstated. Measurement is not about producing numbers. It is about producing numbers that preserve empirical structure. Rasch measurement achieves this by enforcing conjoint additivity between persons and items, yielding a linear logit ratio scale with invariant units. Without this property, numerical outputs may be ordered, averaged, or modeled, but they do not constitute measurement in the representational sense.

This immediately exposes the fatal limitation of multiattribute health instruments. Such instruments fail before measurement can even begin. Measurement presupposes a single attribute. Unidimensionality is not a desirable feature; it is a necessary one. An attribute must be one thing. Vision, mobility, pain, mood, cognition, and self-care are not manifestations of a single latent variable. They are qualitatively distinct phenomena. Aggregating them does not create a higher-order attribute. It creates an index.

No amount of valuation, weighting, or statistical sophistication can overcome this initial failure. If there is no single attribute, there is nothing to measure. Rasch measurement cannot be applied because there is no latent construct to possess. The failure is therefore not technical but ontological. Multiattribute instruments do not fail because Rasch has not been applied. They fail because Rasch cannot be applied.

This is why preference-based systems substitute valuation for measurement. Where attributes cannot be unified empirically, preferences are invoked to impose numerical coherence. Health states are described ordinally, weighted by social judgments, collapsed into a single index, and then treated as if that index represents magnitude. The resulting numbers may be stable, reproducible, and widely accepted, but they do not possess invariant units because nothing in the structure preserves an empirical relation between quantities.

The absence of invariance at the instrument level explains why downstream modeling cannot rescue such systems. If the source numbers do not represent quantities, no modeling framework can restore meaning. Arithmetic applied to non-quantities remains non-quantitative regardless of sophistication. Invariance cannot be introduced after the fact. It must exist at the point of construction.

The concept of invariance, however, extends beyond statistical modeling. At a deeper level, invariance can characterize epistemic systems themselves. Epistemic invariance refers to the persistence of the same conceptual assumptions, exclusions, and numerical permissions across institutional settings. When different systems independently generate the same patterns of reasoning, the same types of claims, and the same absences of constraint, the explanation cannot lie in local error. It must lie in a shared underlying structure of thought.

This is the sense in which structural invariance is used in the Logit Working Papers. The reduced canonical diagnostics do not test respondents, instruments, or datasets. They interrogate knowledge bases. They examine whether certain propositions function as governing rules within a system; whether measurement axioms operate as admissibility conditions for numerical claims, or whether they are absent altogether. When multiple national systems, operating under different legal frameworks, administrative traditions, academic cultures, and policy mandates, exhibit the same endorsement profile, the result is decisive. The convergence cannot be attributed to coincidence, local misunderstanding, or training deficiencies. It reveals a common epistemic inheritance.

Structural invariance, in this context, therefore signals something far stronger than widespread error. It indicates non-possession. Measurement theory does not function as a governing authority within these systems. The axioms that determine when numbers may represent quantities are not recognized as constraints. Where axioms are absent, they cannot be selectively applied. Numerical practice proceeds not by violating rules, but by operating in a space where the rules do not exist.

This is why the reduced canonical diagnostics yield structurally invariant results across countries and instruments. In every case, the propositions that would constrain numerical practice under representational measurement theory—unidimensionality, invariance of units, admissible scale type, and the rule that measurement must precede arithmetic—collapse toward strongly negative logits. That invariance is the central finding. What varies is the *surface polarity* of the false propositions. In most systems, the assumptions that make multiattribute utility arithmetic possible register as strongly positive reinforcement; in others (notably the 15D), the same assumptions can appear with mixed or even uniformly negative polarity, reflecting differences in how the knowledge base narrates, defends, or disguises the same underlying category error. The coherence lies in the shared outcome: arithmetic is authorized without measurement, because measurement axioms do not operate as governing constraints..

In Rasch terms, the pattern is invariant; but what is invariant is not a latent trait possessed by the system. It is a belief structure. The diagnostics do not identify measurement competence distributed unevenly across countries; they identify its systematic absence everywhere. What remains stable is not knowledge of measurement, but the shared acceptance of numerical practice without admissibility. The resulting invariance is therefore not evidence of robustness. It is evidence of epistemic closure. Across national HTA systems, non-measurement itself has become the invariant structure.

Recognizing this distinction is essential. It explains why incremental reform repeatedly fails. Adjusting instruments, expanding response levels, refining valuation techniques, or improving modeling frameworks cannot introduce invariance where none exists. Measurement cannot be retrofitted. It must be foundational. Structural invariance thus becomes the keystone concept linking Rasch measurement theory to the epistemic critique advanced in the Logit Working Papers. Rasch demonstrates what measurement requires. Structural invariance demonstrates what happens when those requirements are absent everywhere at once. Together, they reveal why multiattribute health instruments do not merely fall short of measurement standards, but fail at the first step: the existence of a single attribute capable of being measured at all.

THE REDUCED CANONICAL DIAGNOSTIC

The reduced canonical statement interrogation is designed neither to test competence nor to infer intention. It does not assess methodological sophistication, professional training, or analytical diligence. Its purpose is diagnostic rather than evaluative. The objective is to determine whether an epistemic environment contains the conceptual conditions required for measurement to occur at all.

The interrogation does not ask whether a proposition is persuasive, widely cited, or operationally convenient. It asks whether the knowledge base treats that proposition as an operative constraint governing numerical practice. In other words, it asks which principles function as rules of admissibility and which do not. For this reason, the reduced canonical statement set is constructed to include two qualitatively different classes of propositions.

The first class consists of statements that are true under representational measurement theory. These articulate necessary axioms for quantitative representation. They include requirements such as unidimensionality, the precedence of measurement over arithmetic, the restriction of multiplication to ratio scales, the existence of a true zero, and the requirement of invariant units when latent attributes are quantified. Reinforcement of these propositions indicates possession of measurement knowledge. Their absence indicates that measurement theory does not function as a governing authority within the system.

The second class consists of statements that are known to be false under representational measurement theory, yet which are routinely treated as admissible within contemporary health technology assessment practice. These include propositions permitting negative values on ratio scales, treating preference-based indices as quantitative measures, aggregating multiattribute utilities, or interpreting valuation as a substitute for measurement. These statements do not reflect misunderstanding or error. They articulate the precise assumptions required for multiattribute utility frameworks to operate. This distinction is essential for interpreting the probability–logit results.

The interrogation does not assign correctness empirically. Correctness is defined a priori by the axioms of representational measurement. What the interrogation evaluates is reinforcement: whether the epistemic environment endorses, reproduces, teaches, or operationalizes a given proposition as a governing rule. Accordingly, the logit scale does not encode truth. It encodes possession.

For true propositions, low endorsement probabilities generate strongly negative logits. These values indicate that foundational measurement axioms do not function as operative constraints. Their absence is not episodic or partial. Measurement does not precede arithmetic. Unidimensionality does not restrict aggregation. Ratio scale requirements do not discipline multiplication. These propositions do not operate within the epistemic environment.

For false propositions, high endorsement probabilities generate strongly positive logits. These values do not indicate epistemic success. They indicate normalization. The positive sign reflects acceptance, not validity. This polarity is therefore not contradictory. It is complementary.

The reduced canonical diagnostic is intentionally directional. True propositions are scored in one direction; false propositions in the other. When an epistemic system lacks measurement axioms, it cannot remain neutral. In order for numerical practice to persist, the system must positively endorse propositions that permit arithmetic to proceed without measurement. False propositions must be accepted, not merely tolerated.

This is why the final cluster of false statements consistently exhibits strong positive logits across instruments and countries. These propositions correspond exactly to the conceptual moves that sustain multiattribute utility systems. Without acceptance of these assumptions, QALYs could not be aggregated, negative utilities could not exist, preference scores could not be treated as magnitudes, and simulation outputs could not be interpreted quantitatively. Their reinforcement is therefore constitutive rather than accidental.

The resulting pattern reveals an epistemic inversion. What should function as constraints is absent. What should be prohibited is normalized. Measurement axioms disappear, while their negation becomes routine practice. The logit structure captures this inversion with exceptional clarity.

This is where Rasch invariance becomes decisive. In Rasch measurement, invariance is not a statistical convenience but a defining property of measurement itself. Item difficulty and person ability must be separable. Comparisons must remain stable across populations and across item sets. Measurement, once established, does not depend on context. Where invariance fails, measurement does not exist.

The reduced canonical interrogation mirrors this logic, not at the level of person responses, but at the level of epistemic structure.

Across countries, agencies, journals, and instruments, the same polarity pattern repeatedly appears. True measurement axioms collapse toward strong negative logits. False quantitative assumptions cluster toward strong positive logits. This consistency is not noise. It is structural invariance.

Just as Rasch invariance reveals possession of a latent trait, epistemic invariance reveals possession of a belief structure. In this case, what is possessed is not measurement competence, but its absence. The stability of the pattern demonstrates that the system does not misunderstand measurement differently in different places. It does not misunderstand measurement at all. It does not possess it. This is the defining insight.

The observed pattern is not one of confusion, partial knowledge, or transitional reform. It is one of coherent non-measurement. The epistemic environment is internally consistent. It simply operates under an ontology in which numerical authority derives from consensus, repetition, and institutional endorsement rather than from representational validity.

This explains why nearly identical diagnostic profiles emerge across EQ-5D, HUI, AQoL, and 15D instruments. Differences in descriptive content, valuation method, and national tradition are irrelevant once the epistemic structure is shared. Where measurement axioms are absent everywhere, identical outcomes must follow everywhere.

The polarity of the logits therefore performs a critical interpretive function. Negative values on true propositions reveal non-possession of measurement principles. Positive values on false propositions reveal normalization of false quantitative assumptions. Together, they map the full epistemic architecture within which contemporary health technology assessment operates.

What may appear superficially as contradiction is, in fact, coherence. The system is not oscillating between correct and incorrect ideas. It is consistently endorsing a framework in which valuation replaces measurement and arithmetic precedes representation.

The logit results do not expose marginal error. They expose the governing logic of the enterprise itself.

This is why incremental reform cannot succeed. Adjusting instruments, refining tariffs, or improving modeling sophistication cannot alter an epistemic structure that has inverted admissibility conditions. Measurement cannot be introduced downstream. It must exist at the point where numbers first claim to represent empirical attributes.

The reduced canonical diagnostic therefore functions not as a critique of particular instruments, but as a structural mirror. It reflects back to national HTA systems the logic they already operate under. The mirror does not distort. It clarifies.

Once this structure is recognized, the meaning of the results becomes unavoidable. The problem is not that measurement is poorly implemented. It is that measurement is absent — and its negation has become invariant.

The reduced canonical statement set employed across instrument-level analyses isolates pre-arithmetic conditions. These include propositions concerning unidimensionality, scale-type distinctions, the existence of a true zero, invariance of units, the requirement that measurement precede arithmetic, the distinction between valuation and measurement, and the necessity of Rasch transformation for latent traits. These propositions are not methodological preferences. They are logical conditions. They determine whether numbers may legitimately represent empirical attributes and whether arithmetic operations are admissible at all.

The reduced set deliberately excludes downstream policy claims, aggregation rules, and modeling conventions. Its purpose is to interrogate the boundary at which numbers are first claimed to represent quantity. If measurement fails at this boundary, no subsequent analytical sophistication can repair it. Each instrument-level assessment therefore applies the same reduced statements, the same probability–logit transformation, and the same interpretive logic. Differences in outcome consequently reflect differences in epistemic structure — not differences in analytic method.

FOUR INSTRUMENTS, FOUR COUNTRIES

The instruments examined under this framework originate in four distinct national contexts.

The EQ-5D emerged from European efforts to standardize health-state description and valuation, later becoming embedded in UK HTA through NICE and subsequently diffused across Europe.

The Health Utilities Index was developed in Canada as a multiattribute classification and valuation system, later incorporated into national and international economic evaluations.

The Assessment of Quality of Life was developed in Australia as a locally grounded alternative, later expanded into a family of instruments explicitly positioned for utility measurement and QALY construction.

The Finnish 15D was developed as a national instrument emphasizing descriptive richness, yet ultimately incorporated into valuation-based indexing and economic analysis.

These instruments differ substantially in appearance. They vary in the number of domains, response structures, valuation techniques, and stated design motivations. On superficial inspection, they appear epistemically diverse.

The reduced-item diagnostics reveal otherwise. These instruments differ substantially in appearance. They vary in domains, response formats, valuation procedures, and stated design rationales. Yet none of these differences bear on the conditions required for quantitative representation. The reduced-item diagnostics reveal otherwise.

RESULTS

The reduced canonical statement interrogation produces a highly structured and internally consistent pattern across the four instruments examined: EQ-5D, HUI, AQoL, and the Finnish 15D. While numerical magnitudes differ modestly, the directional structure of endorsement is stable and reveals a clear separation between two classes of propositions: those testing the pre-arithmetic conditions of measurement and those required to sustain valuation-based numerical practice.

TABLE 1 REDUCED ITEM STATEMENT, RESPONSE AND NORMALIZED LOGITS

STATEMENT	RESPONSE 1=TRUE 0=FALSE	NORMALIZED LOGIT (+/- 2.50)			
		EQ-5D	HUI	AQoL	15D
INTERVAL MEASURES LACK A TRUE ZERO	1	-1.40	-1.40	-1.40	-1.40
MEASURES MUST BE UNIDIMENSIONAL	1	-2.20	-2.20	-2.20	-2.20
MULTIPLICATION REQUIRES A RATIO MEASURE	1	-2.20	-2.20	-2.20	-2.20
MEASUREMENT PRECEDES ARITHMETIC	1	-2.20	-2.20	-2.20	-2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	-2.20	-2.50	-2.50	-2.50

TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	-2.50	-2.50	-2.50	-2.50
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FO ASSESSING LATENT TRAIT IMPACT	1	-2.50	-2.50	-2.50	-2.50
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	-2.50	-2.50	-2.50	-2.50
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	-2.50	-2.50	-2.50	-2.50
PREFERENCE BASED UTILITIES CREATE INTERVAL MEASURES	0	-2.50	-2.50	-2.50	+1.75
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	-2.50	-2.50	-2.50	+2.20
MULTIATTRIBUTE HEALTH STATE CLASSIFICATGIONS ARE UNIDIMENSIONAL	0	-2.50	-2.50	-2.50	+2.20
ORDINAL HEALTH STATE DESCRIPTIONSCAN BE TRANSFORMED INTO QUANTITATIVE MAGNITUDE THROUGH PREFERENCE WEIGHTING	0	-2.50	-2.50	-2.50	+1.75
PREFERENCE ALGORITHM SCORING PRODUCES MEASUREMENT VALID NUMERICAL QUANTITIES	0	-2.50	-2.50	-2.50	+1.75

The first nine reduced statements test foundational requirements derived from representational measurement theory. These include the necessity of unidimensionality, the restriction of multiplication to ratio scales, the requirement that measurement precede arithmetic, and the conditions governing latent trait measurement, including Rasch invariance and trait possession.

Across all four instruments, these propositions exhibit uniformly negative logits. In every case, endorsement probabilities are low, and in most instances collapse toward the floor of the scale. This pattern holds without exception for the core axioms of measurement. Regardless of instrument origin, descriptive structure, or valuation method, none of the knowledge bases examined reinforce these propositions as operative constraints governing numerical interpretation.

Notably, this convergence is exact in direction. While small variations appear in magnitude, the sign structure of the first nine items is invariant across instruments. The result is not a mixture of partial endorsement or transitional uncertainty, but a stable pattern in which foundational

measurement requirements do not function as admissibility conditions at the point where numerical outputs are first generated.

The remaining five statements in the reduced set address propositions known to be false under representational measurement theory but necessary for the operation of multiattribute utility systems. These include assumptions concerning the quantitative status of preference-based utilities, the permissibility of negative values on purported ratio scales, the treatment of heterogeneous health state classifications as unidimensional, and the capacity of valuation algorithms to generate quantitative magnitude.

For the EQ-5D, HUI, and AQoL instruments, these final five propositions also exhibit negative logits. In these cases, the knowledge bases do not explicitly affirm the false propositions as theoretical claims. Rather, valuation-based arithmetic proceeds without overt doctrinal articulation. The assumptions required for quantification are not formally defended or rejected; instead, they remain embedded tacitly within routine analytic practice.

The Finnish 15D instrument displays a distinct numerical configuration for this final cluster. While the first nine items align exactly with the other instruments, the last five exhibit strong positive logits. This polarity difference does not alter the overall structural pattern but reflects a difference in the mode through which valuation assumptions are sustained. In the 15D framework, preference weighting and index construction are articulated more explicitly as quantitative transformations rather than operating silently as background conventions.

This difference is limited strictly to the final five items. It does not affect the foundational measurement axioms, all of which remain uniformly negative across all four instruments. The divergence therefore does not indicate epistemic departure at the level of measurement theory. Instead, it reflects variation in how valuation assumptions are expressed—explicitly in the 15D case, implicitly in the others.

Taken together, the reduced canonical results reveal a highly constrained outcome space. Across all instruments, propositions defining the logical conditions for measurement are absent as governing constraints. Where numerical practice continues, it does so through assumptions that either remain tacit or are explicitly affirmed, depending on instrument tradition. The variation observed concerns the visibility of those assumptions, not their function.

The structure of the results therefore isolates a common boundary condition shared by all four systems. The divergence among instruments occurs only after that boundary has already been crossed. Measurement axioms fail uniformly upstream; valuation assumptions determine numerical practice downstream. This separation clarifies the relationship between convergence and variation observed in the table.

The results do not demonstrate inconsistency across instruments. They demonstrate structural alignment at the level that determines whether measurement occurs at all, combined with limited variation in how non-measurement is operationalized once arithmetic is underway.

CONCLUSION

The reduced canonical results reported in this analysis establish a striking and consequential empirical pattern. Across four independently developed national multiattribute instruments—the EQ-5D, HUI, AQoL, and 15D—the same foundational measurement propositions collapse in precisely the same direction. The first nine reduced items, which test the pre-arithmetic conditions of representational measurement, are uniformly negative across all instruments and countries. This invariance is not approximate or impressionistic. It is exact in direction and tightly clustered in magnitude. Regardless of differences in descriptive content, valuation method, institutional history, or national context, none of the knowledge bases examined recognize unidimensionality, invariant units, or the logical precedence of measurement over arithmetic as operative constraints. Measurement theory does not function anywhere as an admissibility condition for numerical claims.

This finding alone is sufficient to rule out explanations based on local misunderstanding, educational gaps, or methodological transition. If measurement failure were contingent, one would expect variation: partial endorsement, mixed signs, or inconsistent collapse. Instead, the pattern is identical across three continents and four decades of instrument development. The absence of measurement axioms is therefore not accidental. It is structural. What the diagnostics reveal is not error but non-possession. Measurement theory is not misapplied or imperfectly understood. It is simply not part of the epistemic grammar within which health technology assessment operates.

The invariance of the first nine items has decisive interpretive consequences. These items do not concern modeling conventions, policy preferences, or analytic taste. They concern the logical conditions under which numbers may represent empirical attributes at all. When these conditions are absent, numerical outputs may still be generated, manipulated, and circulated, but they cannot support quantitative inference. Arithmetic may proceed, but it does so without representational authorization. The diagnostic therefore identifies the point at which HTA departs from measurement science and enters a different epistemic domain entirely—one in which numerical legitimacy derives from institutional repetition rather than from structural correspondence between number and attribute.

The final five items in the reduced set illuminate what occurs once that departure has taken place. These propositions articulate the assumptions required for multiattribute valuation systems to function: that preference-based utilities create interval measures, that negative values may exist on purported ratio scales, that heterogeneous domains may be treated as unidimensional, and that algorithmic scoring can generate quantitative magnitude. These propositions are false under representational measurement theory, yet they are indispensable to QALY construction. Without their acceptance, arithmetic collapses downstream.

It is here that the Finnish 15D diverges numerically from the other instruments, and this divergence is analytically revealing rather than problematic. For the EQ-5D, HUI, and AQoL, all fourteen reduced items exhibit negative logits. The knowledge bases surrounding these instruments do not explicitly affirm false valuation propositions. Instead, valuation is normalized tacitly. Arithmetic proceeds through convention, precedent, and routine use, without overt theoretical declaration. The assumptions required for quantification are not defended; they are simply never questioned.

The Finnish 15D behaves differently. While it exhibits the same collapse on all foundational measurement axioms, its final five items display strong positive logits. This does not indicate superior epistemic status. On the contrary, it indicates explicit doctrinal endorsement of valuation assumptions. The 15D does not merely rely on valuation implicitly; it affirms it openly. Preference weighting is explicitly treated as quantitative transformation. Negative values are openly permitted. Multiattribute health state descriptions are explicitly treated as if they were unidimensional magnitudes.

This difference matters, but not in the way that conventional methodological debates would suggest. The Finnish instrument does not escape measurement failure. It formalizes it. Where the other systems rely on silent normalization, the 15D codifies the substitution of valuation for measurement as an explicit theoretical position. The result is not epistemic progress, but epistemic transparency. Finland does not possess measurement theory, but it is unusually clear about what replaces it.

This distinction resolves what initially appeared puzzling about the sign reversal in the final five logits. The divergence is not inconsistency. It is difference in visibility. All four systems abandon measurement at the same point. What differs is whether the replacement ontology is implicit or explicit. In the EQ-5D, HUI, and AqOL traditions, valuation operates as an unspoken assumption embedded in practice. In the 15D tradition, valuation is articulated as principle. The outcome is numerically different but epistemically equivalent.

The broader implication is sobering. Health technology assessment has not developed as a measurement science that later adopted utilities for convenience. It developed as a valuation enterprise from its inception. Numerical outputs were never required to satisfy representational conditions because their authority did not originate in measurement. It originated in preference. Arithmetic was adopted not because quantities existed, but because decisions demanded numbers. Once arithmetic became institutionalized, the question of whether those numbers represented anything ceased to arise.

This explains the extraordinary resilience of the HTA numerical framework. It is not sustained by theoretical argument. It is sustained by closure. Developers can point to use. Users can point to guidelines. Agencies can point to precedent. Educators can point to curricula. At no point does the system encounter a requirement to justify quantification itself. The reduced canonical diagnostics expose this closure with unusual clarity. They show that the system is internally coherent precisely because it does not contain measurement axioms that could destabilize it.

The results also clarify why reform efforts have repeatedly failed. Proposals to improve sensitivity, refine valuation protocols, or enhance modeling sophistication all operate downstream of the pre-arithmetic boundary. They assume that quantities already exist. But where measurement has not occurred, no refinement can create it. Unidimensionality cannot be introduced into a multiattribute construct by recalibration. Invariant units cannot be produced by preference elicitation. A true zero cannot be manufactured through algorithmic anchoring. These are not technical limitations. They are logical impossibilities.

The convergence observed across instruments and countries therefore removes discretion from the debate. It is no longer plausible to argue that particular agencies made poor choices, that particular instruments were badly designed, or that particular training programs were incomplete. The failure is architectural. The epistemic system as a whole operates without measurement as an admissibility condition. That is why the same reduced diagnostic profile appears everywhere measurement is tested.

The consequences for inference are profound. When numerical variation cannot be shown to correspond to variation in an underlying attribute, claims about magnitude, improvement, or comparative effectiveness lose their empirical grounding. Differences may reflect changes in valuation structure rather than changes in health itself. Precision becomes computational rather than empirical. Confidence intervals surround quantities that do not exist.

None of this implies that descriptive classification or preference research lacks value. These activities can inform deliberation, priority setting, and qualitative comparison. The error arises only when valuation outputs are elevated to quantitative status and treated as measures of magnitude. Once that elevation occurs, arithmetic acquires an authority it cannot logically sustain.

The reduced canonical results therefore do not merely critique particular instruments. They identify the boundary condition of contemporary HTA. The field has achieved remarkable administrative coordination and methodological elaboration, but it has done so by consolidating a system of numerical storytelling rather than constructing a measurement science. The invariance observed across EQ-5D, HUI, AQoL, and 15D demonstrates that this is not an accidental outcome. It is the defining feature of the enterprise.

Recognizing this does not require hostility toward existing institutions. It requires clarity. Either measurement is required for quantitative claims, or it is not. If it is required, then the current architecture cannot supply it. If it is not required, then HTA must acknowledge that its numbers do not represent quantities and abandon the language of magnitude, precision, and quantitative inference. There is no stable intermediate position.

The reduced canonical diagnostics make that choice unavoidable. They do not argue for reform. They reveal necessity. They show that health technology assessment has converged internationally not on measurement, but on valuation endowed with arithmetic authority. Until that distinction is confronted explicitly, numerical practice will continue to appear scientific while remaining epistemically ungrounded.

ACKNOWLEDGEMENT