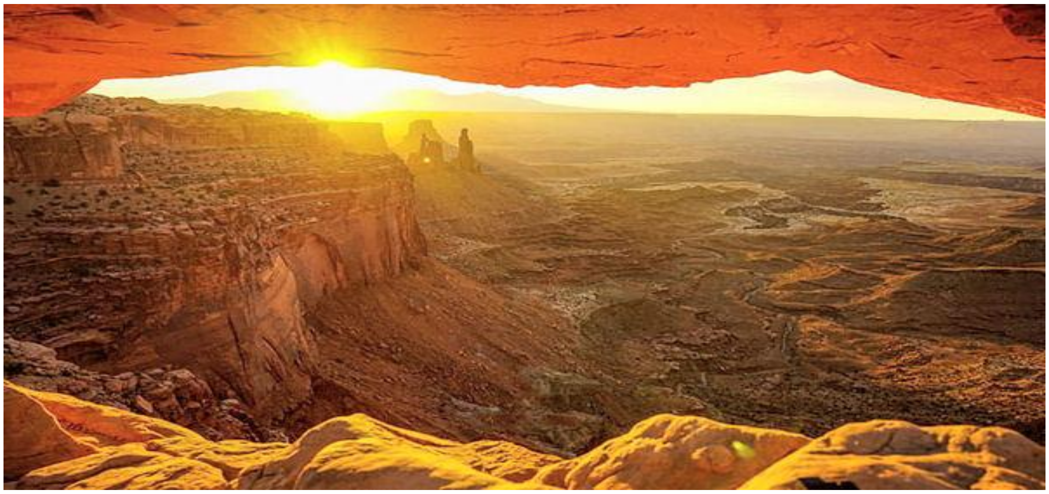
MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION



REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

AUSTRALIA: DECONSTRUCTING THE EPISTEMIC KNOWLEDGE BASE OF THE AUSTRALIAN AQoL

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

LOGIT WORKING PAPER No 37 FEBRUARY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

The Assessment of Quality of Life (AQoL) instruments constitute the principal Australian-developed family of preference-based health-related quality of life measures and are widely used in clinical research, population studies, and health technology assessment. Across successive versions, including the AQoL-4D, AQoL-6D, and AQoL-8D, the instruments have expanded in descriptive scope, particularly in psychosocial and mental health domains, and have been explicitly positioned to support utility estimation and quality-adjusted life year (QALY) construction. Despite their prominence, the measurement status of AQoL numerical outputs has not been subjected to systematic interrogation against the axioms of representational measurement theory.

This working paper examines the AQoL instrument family not as a technical artifact in isolation, but as an epistemic object embedded within a distributed user system. Numerical authority, it is argued, does not arise from instrument construction alone, but from repeated use within an environment that treats outputs as quantities. The object of analysis is therefore the epistemic knowledge base that authorizes AQoL utilities to be averaged, differenced, and multiplied by time, rather than the intentions of developers or the correctness of individual applications.

To evaluate this knowledge base, the paper applies a reduced canonical diagnostic derived from representational measurement theory and Rasch principles. The diagnostic consists of propositions that test whether foundational conditions for quantitative measurement such as unidimensionality, invariant units, admissible scale type, the existence of a true zero, and the logical precedence of measurement over arithmetic which function as operative constraints within the AQoL epistemic environment. Endorsement probabilities are transformed into normalized logits to reveal structural patterns of possession and absence.

The results demonstrate a stable and internally coherent epistemic profile. Propositions expressing necessary conditions for measurement consistently exhibit low endorsement, indicating that measurement axioms do not function as governing constraints. Conversely, propositions known to be false under measurement theory, such as the treatment of preference-based multiattribute indices as quantitative magnitudes and the permissibility of negative values on purported ratio scales exhibit strong positive endorsement. This polarity reveals not confusion or partial misunderstanding, but structural non-possession of measurement theory.

The findings indicate that AQoL instruments do not produce quantities in the representational sense. Their numerical outputs express valuation, not magnitude. Expansion of descriptive content across instrument versions does not alter this status, as refinement within a multiattribute valuation framework cannot establish measurement properties post hoc.

The paper concludes that AQoL functions coherently as a preference valuation system but cannot support quantitative claims concerning magnitude, change, or comparative therapeutic impact. This conclusion aligns with parallel findings for the EQ-5D, HUI, and 15D instruments, demonstrating epistemic structural invariance across national multiattribute frameworks. Restoring measurement as an admissibility condition for quantitative claims is therefore

essential if health technology assessment is to move beyond numerical convention toward empirically meaningful knowledge.

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 Assessment of Quality of Life (AQoL) instruments constitute a family of multiattribute health-related quality of life measures developed in Australia under the leadership of the Centre for Health Economics at Monash University [1]. The AQoL program was initiated in the late 1990s with the explicit aim of providing a preference-based alternative to the EQ-5D suitable for utility estimation and QALY construction, while offering greater descriptive sensitivity, particularly in psychosocial domains.

The original AQoL (AQoL-4D) comprised four dimensions—independent living, social relationships, physical senses, and psychological wellbeing—represented through multiple items and aggregated using population preference weights derived from time trade-off methods. Subsequent versions expanded the descriptive system substantially. AQoL-6D and AQoL-8D introduced additional domains, with the AQoL-8D incorporating extensive mental health and social functioning content intended to improve responsiveness in mental health applications.

Across all versions, AQoL instruments follow the same conceptual architecture. Ordinal responses across multiple heterogeneous domains are combined through scoring algorithms based on preference elicitation to produce a single utility index anchored on death and permitting negative values. These utilities are explicitly intended for use in cost-utility analysis and QALY construction. The AQoL family has been widely applied in Australian clinical studies, population surveys, and health economic evaluations, and is frequently presented as a domestically grounded alternative to European utility instruments while remaining fully compatible with standard HTA frameworks.

This deconstruction of the epistemic basis of the AQoL instrument family 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. What persists instead is numerical storytelling.

The object of analysis is therefore not the AQoL family of instruments 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 AQoL outputs function as if they were quantitative measures. This paper refers to that system as the user epistemic system; a distributed framework of practice that authorizes numerical use through repetition and acceptance, but does not itself confer measurement legitimacy.

The user epistemic system consists of the collective practices through which AQoL values are interpreted as meaningful numbers. It includes researchers who analyze AQoL data, reviewers who evaluate manuscripts, editors who publish results, health technology assessment agencies that accept AQoL 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 AQoL utilities satisfy the axioms required for quantification. Authority arises instead 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 AQoL 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 growth in the number of expanded and “add-on” versions of the AQoL exemplifies this dynamic. Expansion of descriptive content 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 AQoL 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 AQoL 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 AQoL 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 AQoL INSTRUMENTS 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 AQoL system. This cannot be limited to the original development papers produced by the Monash research group, nor can it be confined to formal descriptions of the instrument family. 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 AQoL knowledge base must therefore be understood as an epistemic corpus.

This corpus includes the foundational publications introducing the original AQoL-4D and the subsequent development of the AQoL-6D and AQoL-8D instruments. These texts define a multiattribute classification system covering domains such as independent living, social relationships, mental health, coping, self-worth, pain, and sensory functioning. Each domain is expressed through ordered categorical responses and combined through preference-based scoring algorithms. From the outset, the AQoL instruments were framed not as direct measures of health, but as systems for describing health states to be valued through population preferences. This framing is decisive. It establishes valuation, rather than measurement, as the conceptual foundation of the instrument family.

However, the authority of AQoL 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. Across Australian clinical studies, population health surveys, mental health evaluations, and economic submissions, AQoL utilities are routinely reported as outcomes. Means are calculated, changes interpreted, and differences compared across treatment 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 bodies constitute a second and particularly influential layer of the AQoL knowledge base. Within Australia, AQoL utilities have been widely accepted in economic evaluations submitted to the Pharmaceutical Benefits Advisory Committee (PBAC). This acceptance confers institutional authority. It does not arise from demonstration that AQoL values satisfy the axioms of representational measurement, but from their compatibility with established cost-utility frameworks. Once incorporated into submissions and precedent decisions, the numerical status of AQoL utilities becomes administratively secured rather than theoretically established.

Methodological guidance and submission conventions further reinforce this authority. Economic evaluation manuals, PBAC technical guidance documents, and consultant templates routinely treat AQoL utilities as interchangeable numerical inputs. Utilities are abstracted from their descriptive origins and incorporated into standardized analytic workflows. At this stage, the instrument ceases to function as an epistemic object and becomes part of analytic infrastructure. The question of what kind of numbers AQoL produces is displaced by the assumption that numbers are required.

Education plays a central role in reproducing this knowledge base. In Australian health economics training programs and professional short courses, students are taught how to apply AQoL utilities in modeling exercises. They learn to calculate QALYs, to estimate incremental cost-effectiveness ratios, and to interpret threshold results. 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 AQoL knowledge base extends further through analytic infrastructure. Software packages, spreadsheet models, and consultancy templates embed AQoL scoring algorithms and Australian value sets. Once encoded, conceptual assumptions become invisible. Users interact with numerical outputs without encountering the premises that authorize their treatment as quantities. In this way, epistemic commitment is no longer expressed through argument or citation, but through automation.

The expansion of the AQoL family illustrates how the knowledge base absorbs modification without altering its foundations. The progression from AQoL-4D to AQoL-6D and AQoL-8D is commonly presented as methodological progress, improving sensitivity, particularly in psychosocial domains. Yet this elaboration does not alter the underlying epistemic architecture. The instruments remain multiattribute. Valuation remains preference-based. Utilities continue to permit negative values. Arithmetic compatibility with QALY construction remains assumed rather than demonstrated. The proliferation of versions therefore represents elaboration within a stable belief system rather than epistemic transformation.

Crucially, the AQoL knowledge base is not unified by explicit theoretical agreement. There is no authoritative text asserting that AQoL utilities satisfy the axioms of representational measurement. 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 AQoL system. Developers can point to widespread application. Users can point to accepted methodological practice. 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 AQoL 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.

III. INTERROGATING THE AQoL KNOWLEDGE BASE

If the numerical authority of the AQoL 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 AQoL scoring correctly, nor whether its developers intended to create a measure. The relevant question is whether the knowledge base that authorizes the numerical use of AQoL 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 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 quality of life, functioning, or psychological wellbeing 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 such as AQoL 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 AQoL 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 AQoL 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 AQoL system, this framework allows the analysis to move beyond debates over descriptive sensitivity, domain coverage, valuation protocol, or empirical responsiveness. Such debates presuppose that the instrument already produces quantities. The present interrogation suspends that presupposition. It asks instead whether the knowledge base that sustains AQoL 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 AQoL system.

IV RESULTS AND DISCUSSION

The reduced canonical diagnostic applied to the Australian Assessment of Quality of Life (AQoL) instrument (Table 1) yields a result that is not ambiguous, transitional, or internally conflicted. Across all fourteen propositions, endorsement probabilities cluster decisively toward the lower bound of the scale, producing uniformly negative normalized logits. No proposition approaches neutrality, and none registers positive reinforcement. The resulting profile is internally coherent and structurally stable. It does not indicate partial misunderstanding, methodological uncertainty, or competing epistemic commitments. Rather, it reveals the systematic absence of representational measurement theory as a governing authority within the epistemic environment that sustains AQoL's numerical use.

TABLE 1

REDUCED ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED LOGITS AUSTRALIA AQoL INSTRUMENTS

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

This outcome is especially instructive because AQoL was developed explicitly as a national alternative to the EQ-5D, with the stated intention of improving conceptual richness, sensitivity, and psychometric performance. Unlike the EQ-5D, AQoL incorporates a larger number of attributes, deeper item hierarchies, and complex weighting structures intended to better reflect psychosocial dimensions of health. Yet the diagnostic demonstrates that such elaboration does not alter the epistemic status of the resulting numbers. Increased descriptive sophistication does not translate into measurement.

The first group of propositions addresses the most basic 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 concept is not entirely absent from the AQoL knowledge base, it does not function as a constraint on numerical interpretation. AQoL utilities are routinely treated as if zero represented the absence of health, even though no empirical structure is provided to justify such an interpretation.

This ambiguity is not trivial. A true zero is not a semantic convenience but a defining property of ratio measurement. Without it, multiplication and division are undefined. Yet AQoL utilities are routinely multiplied by time, averaged across populations, and interpreted as magnitudes of gain or loss. The diagnostic indicates that these operations proceed independently of scale-type recognition. Zero functions operationally rather than representationally. It is accepted because it appears numerically useful, not because it corresponds to an empirical null state.

The requirement of unidimensionality collapses even more decisively. The proposition that measures must be unidimensional receives an endorsement probability of 0.10, producing a logit of −2.20. This indicates near-total absence of reinforcement. Unidimensionality does not operate as a governing rule within the AQoL epistemic system. This absence is decisive because AQoL is explicitly multiattribute by design, combining independent domains such as independent living, social relationships, mental health, coping, pain, and sensory function.

From the perspective of representational measurement theory, such heterogeneity is fatal to measurement. Measurement requires that numerical variation correspond to variation along a single attribute. Where multiple attributes are combined, the resulting index does not measure anything in particular. It reflects weighting conventions rather than empirical magnitude. The diagnostic demonstrates that this distinction does not constrain practice within the AQoL environment. Multiattribute aggregation is treated as compatible with measurement, despite violating its most basic requirement.

Closely related is the proposition that measurement must precede arithmetic. This item also registers at 0.10 (−2.20). The implication is direct: arithmetic operations are undertaken without prior establishment of measurement properties. In the AQoL system, numbers are manipulated first and justified later, if at all. Arithmetic becomes the mechanism through which numerical authority is constructed rather than the operation whose admissibility must be justified. 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 AQoL environment, this conditionality is absent. Arithmetic becomes routine practice rather than logically constrained inference. Numbers are treated as quantities because they behave numerically, not because they represent quantities.

This inversion is reinforced by the similarly low endorsement of the proposition that multiplication requires a ratio measure. With an endorsement probability of 0.10, scale type does not function as a constraint on analytic practice and the application of arithmetic.. Whether a quantity can be meaningfully multiplied is treated as a modeling decision rather than a measurement question. This inversion lies at the heart of utility-based economic evaluation.

The most decisive results emerge when the diagnostic turns to latent trait measurement. 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 AQoL knowledge base does not recognize Rasch transformation as a necessary condition for converting ordinal responses into interval measures. It does not recognize the Rasch logit ratio scale as the only defensible basis for latent trait quantification. Nor does it recognize the equivalence between Rasch axioms and representational measurement theory.

These findings are not incidental. Latent constructs such as quality of life, functioning, or well-being are not directly observable. Their measurement requires a model capable of producing invariant units across persons and items. Without such a model, numerical scores remain ordinal regardless of how many items are included or how sophisticated the scoring algorithm appears. AQoL does not include such a transformation. Instead, it relies on preference elicitation to assign numerical values to health states. Preference, however, expresses desirability, not magnitude. It provides ordering, not quantity. It cannot generate invariant units. The collapse of Rasch-related items therefore reflects not oversight but incompatibility. Rasch measurement requires unidimensionality and invariance, both of which are explicitly violated by AQoL's multiattribute architecture and external valuation framework.

The proposition that the outcome of interest for latent traits is possession of that trait also collapses to the floor. With an endorsement probability of 0.05, the AQoL epistemic system does not conceptualize quality of life as an attribute possessed by individuals. Instead, individuals are mapped onto health-state profiles that are valued externally by population preferences. The resulting number expresses how desirable a state is, not how much of an attribute the individual possesses. This distinction is fundamental. Measurement concerns attributes of entities. Valuation concerns judgments of observers. AQoL operates entirely within the latter domain. Treating its outputs as measures of individual health therefore entails a category error that the epistemic system does not recognize.

The remaining false propositions reinforce this conclusion. The proposition that preference-based utilities create interval measures collapses completely. Despite decades of use, the knowledge base does not articulate or defend the claim that preference elicitation yields metric quantities. Interval properties are assumed by convention, not established by theory. Similarly, the rejection of the proposition that ratio measures can have negative values is complete. AQoL permits negative utilities, allowing states to be valued as worse than dead. Yet negative values are incompatible with ratio measurement, which requires a true zero representing absence of the attribute. This contradiction is absorbed without reconciliation. The epistemic system tolerates it because valuation allows negativity, even though measurement does not.

The proposition that multiattribute health-state classifications are unidimensional also collapses entirely. This result is unsurprising, yet its implications are profound. AQoL explicitly combines heterogeneous domains that share no empirical unit. Aggregation is achieved through weighting, not concatenation of an attribute. The resulting index therefore lacks additive structure in the representational sense. Nonetheless, arithmetic proceeds as if additivity existed.

The final propositions confirm the same pattern. Ordinal health-state descriptions are treated as transformable into magnitude through preference weighting, and algorithmic scoring is assumed to confer quantitative legitimacy. Both assumptions collapse under interrogation. Algorithms can assign numbers. They cannot establish representation.

Viewed as a whole, the endorsement profile is perfectly coherent. There are no contradictions, no transitional patterns, and no partial recognition of axioms. All propositions that would constrain numerical use are absent. All assumptions required to treat AQoL outputs as measures are unsupported. The profile is stable because the epistemic structure is stable. This uniformity is diagnostically crucial. It demonstrates that AQoL's failure is not due to implementation error, inadequate training, or misinterpretation by users. It reflects structural non-possession. Measurement theory does not function as an admissibility condition within the AQoL knowledge base. Where axioms are not part of disciplinary grammar, they cannot be selectively applied.

Importantly, this result should not be interpreted as a critique of individual researchers or developers. The diagnostic does not assess belief, competence, or intention. It assesses possession. The principles tested do not operate as governing rules within the system. Even technically sophisticated analysts cannot invoke axioms that the system does not recognize.

The AQoL case therefore reinforces the broader pattern observed across EQ-5D, HUI, and 15D. Despite national differences, descriptive variation, and valuation diversity, the epistemic outcome is invariant. Preference-based multiattribute instruments converge on the same non-measurement structure. In other words, AQoL does not fail because it is Australian, because it is complex, or because it is insufficiently refined. It fails because it is a valuation system treated as a measurement instrument. No amount of refinement can alter that category.

As a descriptive framework, AQoL may retain limited value. It can structure health-state reporting and support preference research. What it cannot do is support quantitative inference about magnitude, change, or comparative effect. Those claims require measurement properties that AQoL does not and cannot possess. The reduced canonical diagnostic therefore confirms a central conclusion of the Logit Working Papers series. The problem is not one of instrument quality. It is epistemic substitution. Valuation has been mistaken for measurement, numerical form for quantitative meaning, and arithmetic convenience for representational legitimacy. Until measurement axioms are restored as admissibility conditions for numerical claims, instruments such as AQoL will continue to generate numbers that look quantitative but are not. What persists is not imperfect measurement, but numerical storytelling.

V EPISTEMIC DISASTER: IMPLICATIONS FOR PBAC DECISION MAKING

The epistemic failure exposed by the reduced canonical interrogation of the AQoL instruments has direct and troubling implications for PBAC decision making. These implications do not arise because AQoL is uniquely flawed. They arise because AQoL exemplifies, in its clearest

Australian form, the deeper structural error that underpins all multiattribute utility instruments used within the PBAC framework.

PBAC's economic evaluation process rests on the assumption that utilities quantify health-related quality of life and that QALYs represent meaningful differences in health outcomes. Yet the reduced-item analysis demonstrates that the AQoL system does not satisfy the axioms required for measurement. It does not establish unidimensionality. It does not generate invariant units. It does not possess a true zero. Its numerical outputs therefore cannot support arithmetic operations such as multiplication by time, comparison of magnitudes, or aggregation across individuals. This is not a technical weakness. It is a categorical failure. What AQoL produces are preference scores, not measures. Their numerical form derives from valuation conventions, not from representational structure. Consequently, when PBAC accepts AQoL utilities as quantitative inputs, it is not evaluating measured treatment effects. It is evaluating the arithmetic consequences of a valuation algorithm.

One response might be to argue that AQoL is merely one of several instruments available and that sponsors may instead submit utilities derived from EQ-5D, HUI, SF-6D, or other multiattribute systems. This response, however, offers no epistemic relief. The objections identified here do not apply uniquely to AQoL. They apply to all preference-based multiattribute instruments without exception. Regardless of the instrument chosen, the same structural conditions hold. Health states are described ordinally across multiple attributes. Preferences are elicited over hypothetical states. Algorithmic weights are applied. The resulting index is treated as if it represented quantitative magnitude. Changing the descriptive system or valuation protocol does not alter this logic. It merely substitutes one valuation convention for another.

From a measurement perspective, the instruments differ in appearance but not in kind. The epistemic failure is invariant. PBAC therefore faces a deeper problem than instrument selection. If AQoL utilities are rejected on measurement grounds, then the same objections must also apply to every alternative utility instrument currently admissible within PBAC submissions. There is no safe substitution available within the existing framework. The problem lies not with the choice of instrument, but with the assumption that valuation can substitute for measurement.

At the same time, PBAC guidelines insist, correctly, that utilities used in submissions should reflect the preferences of the Australian population. This requirement introduces a further contradiction. Even if Australian-specific valuation studies are used, population relevance cannot compensate for the absence of measurement. Preferences may be locally appropriate, culturally grounded, and methodologically sophisticated, yet still fail to generate quantities. Representational validity does not depend on whose preferences are measured, but on whether what is produced constitutes a measure at all.

An Australian preference weight attached to a non-measure remains a non-measure. This creates an epistemic impasse. PBAC requires Australian valuation while simultaneously presuming that valuation produces quantitative health outcomes. Yet the reduced canonical diagnostics demonstrate that valuation, regardless of population source, does not satisfy the admissibility conditions required for arithmetic inference. The problem therefore cannot be resolved by better surveys, improved elicitation techniques, or more representative samples.

What is at stake is not external validity but ontological legitimacy. As long as PBAC continues to treat preference-weighted multiattribute indices as measures, its quantitative decisions remain logically unsupported. Cost-utility ratios may appear precise, internally consistent, and statistically refined, yet they lack the one property required for scientific inference: lawful measurement. This does not imply that descriptive quality-of-life information or preference research lacks value. Such information can inform deliberation, contextual understanding, and patient experience. But it cannot legitimately function as a quantitative substrate for arithmetic decision rules.

The uncomfortable implication is unavoidable. PBAC's decision framework depends on numerical claims that cannot be defended under representational measurement theory. This is not a problem that can be fixed through refinement. It is not a problem of implementation. It is a problem of foundations. Until PBAC confronts the distinction between valuation and measurement and recognizes that population relevance cannot convert one into the other its quantitative decision-making process will remain an exercise in numerical storytelling rather than scientific evaluation.

VI CONCLUSIONS

The analysis presented in this working paper demonstrates that the AQoL family of instruments does not fail measurement criteria in a partial, ambiguous, or transitional manner. 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 authorizes the numerical use of AQoL 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 AQoL epistemic environment. Where axioms are absent, they cannot constrain practice. Numerical operations may proceed, but they do so without representational authorization.

The expansion of the AQoL family over time does not alter this condition. The transition from AQoL-4D to AQoL-6D and AQoL-8D increases descriptive breadth and domain coverage, particularly in psychosocial and mental health domains. Yet increased descriptive richness does not establish measurement. Refinement within an ordinal and multiattribute framework does not generate invariant units, does not resolve attribute heterogeneity, and does not create a true zero. The epistemic architecture remains unchanged. What appears as methodological advancement is, in fact, elaboration within a belief system that presupposes quantification rather than establishing it.

The reduced diagnostic makes clear that the AQoL 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-based preference elicitation. These values express relative 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 reveal inconsistency or conceptual confusion within the AQoL 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 AQoL outputs as quantities remain unsupported. This coherence is itself diagnostic. It indicates not misunderstanding, 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 expansion of descriptive systems. Measurement cannot be retrofitted. It must be established at the point of construction. As long as AQoL 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 execution, but in epistemic role. The AQoL instruments function exactly as designed: as preference-based valuation systems intended to support QALY construction. The error arises only when outputs from such systems are elevated to quantitative status and treated as measures of magnitude. Valuation is not measurement, and numerical convenience cannot substitute for representational validity.

This conclusion aligns directly with reduced-item interrogations of the EQ-5D, HUI, and 15D instrument families. Despite differences in descriptive content, valuation protocols, and national origins, the epistemic structure is invariant. Each system occupies the same conceptual position: a valuation framework embedded within a knowledge base that treats valuation as measurement in order to sustain arithmetic. Recognizing this does not require abandoning descriptive classification or preference research. It requires epistemic realignment. Instruments such as AQoL may retain value as classificatory or descriptive tools, but only if their numerical outputs are demoted from quantitative status. Claims about magnitude, change, and comparative effect must be restricted to what such instruments 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 AQoL, but a foundation for rethinking what it means to measure health at all. Restoring measurement discipline does not entail abandoning quantitative evaluation. It requires reinstating the logical conditions under which numbers may legitimately represent empirical attributes. Only through such discipline can health technology assessment move beyond 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] Centre for Health Economics, Monash University — AQoL Group angelo.iezzi@monash.edu