

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



**REPRESENTATIONAL MEASUREMENT FAILURE IN  
HEALTH TECHNOLOGY ASSESSMENT**

**AUSTRALIA: CLOSED EPISTEMIC REPRODUCTION  
WITH FIVE AUSTRALIAN HTA ACADEMIC  
KNOWLEDGE BASES SUSTAINING MEASUREMENT  
INVERSION**

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

**LOGIT WORKING PAPER No 810 APRIL 2026**

**[www.maimonresearch.com](http://www.maimonresearch.com)**

**Tucson AZ**

## ABSTRACT

*This study examines the structure of health technology assessment (HTA) knowledge bases within Australian academic research centers through the lens of the philosophy of science and representational measurement. Using a standardized 24-item canonical diagnostic, five university-based HTA knowledge domains were interrogated to assess whether they recognize the axioms required for valid measurement and the conditions necessary for falsification. Each statement was evaluated within the context of the knowledge base, with responses expressed as categorical probabilities and transformed into normalized logits to enable comparison across domains.*

*The results demonstrate a consistent and reproducible pattern. Statements that are true under the axioms of representational measurement, such as the requirement for unidimensionality, dimensional homogeneity, and the precedence of measurement over arithmetic are at best weakly endorsed, with probabilities typically between 0.10 and 0.25 and corresponding negative logits. In contrast, statements that are false under these axioms but embedded within HTA practice such as the treatment of QALYs as ratio measures and the aggregation of utility-based constructs are strongly endorsed, with probabilities between 0.80 and 0.90 and positive logits. This pattern mirrors findings from over 100 global HTA knowledge bases, indicating that the observed structure is not jurisdiction-specific but reproducible across independent domains.*

*A distinctive feature of the Australian results is the partial endorsement of scientific principles such as falsifiability, which does not extend to the core constructs of HTA. This disconnect supports the interpretation of HTA as a system of closed epistemic reproduction, in which methods and assumptions are transmitted and reinforced without engagement with the conditions required for their validation.*

*The study concludes that HTA, as currently practiced in Australia, operates outside the framework of scientific measurement. Its outputs are numerical but not measurable, and its core constructs are not subject to falsification. Addressing this requires a transition to measurement-based evaluation grounded in the axioms of representational measurement and the principles of scientific inquiry.*

### **AUSTRALIA RESEARCH CENTER KNOWLEDGE BASES**

**Centre for Health Economics Research and Evaluation (CHERE), University of Technology Sydney**

**Centre for Health Economics (CHE) Monash University**

**Health Economics Group (HEG) University of Adelaide**

**School of Population and Global Health (MSPGH) University of Melbourne**

**Health Economics and Research Modelling Unit (HERMU) University of Queensland**

## **INTRODUCTION: MEASUREMENT, SCIENCE, AND THE PERSISTENCE OF HTA PRACTICE**

Health technology assessment (HTA) presents itself as a quantitative science. Its claims are expressed in numerical form, its methods rely on formal models, and its outputs are used to inform decisions regarding the allocation of healthcare resources. This appearance of scientific rigor rests on an implicit assumption: that the quantities manipulated within HTA such as utilities, QALYs, and cost-effectiveness ratios possess the properties required for measurement. Without this assumption, the use of arithmetic loses its meaning. Numbers may be generated, but they do not represent measurable attributes.

The philosophy of science provides a clear framework for evaluating such claims. Since the scientific revolution, the legitimacy of quantitative inquiry has depended on two conditions: the establishment of measurement and the possibility of falsification. Measurement ensures that attributes are represented on scales with defined properties, allowing for meaningful comparison and transformation. Falsification ensures that claims can be subjected to empirical test and, if necessary, rejected. Together, these conditions support the evolution of objective knowledge. Where they are absent, numerical reasoning may persist, but it does so outside the boundaries of science.

The present study examines HTA within this framework. Using a standardized 24-item canonical diagnostic grounded in the axioms of representational measurement, a series of knowledge bases associated with Australian academic HTA research centers have been interrogated. Each knowledge base is treated as a defined domain, encompassing its published outputs, methodological conventions, and implicit assumptions. The interrogation evaluates whether propositions that are true under the axioms of measurement are recognized, and whether propositions that violate those axioms are nevertheless endorsed. Responses are expressed as categorical probabilities and transformed into normalized logits, allowing for comparison across statements and across domains.

The purpose of this paper is not only to present the results of these interrogations, but to address a deeper question: why does a consistent pattern emerge in which the conditions required for measurement are not recognized, while incompatible constructs are reinforced? The findings demonstrate that this pattern is not incidental. It is stable across institutions and reproducible across jurisdictions. This raises a philosophical problem. If HTA operates with numerical constructs that do not meet the requirements of measurement, and if these constructs are not subjected to falsification, then on what basis does the framework persist?

The argument developed in this paper is that HTA represents a case of closed epistemic reproduction: a system in which methods and constructs are transmitted, reinforced, and repeated without engagement with the conditions required for their validation. Understanding this persistence requires moving beyond technical critique to a consideration of how knowledge systems are sustained, how they resist challenge, and how they can be replaced.

## **FROM INDIVIDUAL KNOWLEDGE BASE INTERROGATION TO CONSENSUS**

The starting point for this analysis is a series of individual interrogations of Australian HTA academic knowledge bases. Each interrogation was conducted independently, applying a standardized 24-item canonical diagnostic grounded in the axioms of representational measurement. The purpose of these interrogations was to determine, within each domain, whether the conditions required for valid measurement were recognized and whether propositions incompatible with those conditions were nevertheless endorsed. At this stage, no assumption was made regarding the relationship between knowledge bases. Each was treated as a separate and self-contained domain.

The results of these individual interrogations, taken in isolation, were already informative. In each case, statements that are true under the axioms of measurement—such as the requirement for unidimensionality, the necessity of dimensional homogeneity, and the principle that measurement must precede arithmetic—were weakly endorsed. Conversely, statements that are false under those same axioms, but embedded within HTA practice—such as the treatment of QALYs as ratio measures and the aggregation of utility-based constructs—were strongly endorsed. This pattern was observed consistently within each knowledge base. However, at the level of individual analysis, the interpretation remained open. It was possible to attribute the results to local factors: institutional emphasis, methodological preferences, or variations in training.

The transition to synthesis removes this ambiguity. When the results from multiple knowledge bases are brought together and examined in terms of probability and logit ranges, a different picture emerges. The pattern observed within each domain is reproduced across domains, with minimal variation in both magnitude and direction. Statements that are true under measurement theory occupy a consistently negative range of logits; statements that are false occupy a consistently positive range. There is no instance in which a knowledge base reverses this pattern. The convergence is not approximate; it is structural.

This convergence justifies the introduction of the concept of consensus. It is important to distinguish the meaning of consensus in this context from its conventional use. This is not consensus in the sense of agreement among researchers or institutions. There is no evidence of coordinated positions or explicit alignment. Rather, consensus refers to the uniformity of knowledge base structure when subjected to a common interrogation. Independent domains, operating under different institutional arrangements and pursuing different research agendas, produce the same pattern of endorsement across the same set of statements. The agreement is implicit, not declared.

The strength of this consensus lies in its reproducibility. The use of a fixed diagnostic instrument, constrained probability categories, and a standardized transformation to normalized logits ensures that comparisons across knowledge bases are valid. Differences in results can therefore be attributed to differences in the knowledge base, not to variation in method. When such differences fail to appear—when the same pattern is observed repeatedly—the conclusion is that the underlying structure is shared.

This shared structure is not limited to Australia. The same pattern has been documented in Canadian HTA knowledge bases, where a similar set of interrogations produced comparable ranges of probabilities and logits. Despite differences in policy environments, institutional configurations, and academic traditions, the direction and magnitude of endorsement remain aligned. This cross-jurisdictional consistency strengthens the interpretation of consensus. It indicates that the pattern is not contingent on local conditions, but reflects a more general property of HTA as a field.

At this point, the focus of the analysis shifts. The question is no longer whether individual knowledge bases recognize the conditions required for measurement. That question has been answered in the negative. The question is whether the absence of recognition is variable or systematic. The results show that it is systematic. The same propositions are weakly endorsed across all domains; the same incompatible constructs are strongly reinforced. The pattern is stable, reproducible, and resistant to variation.

This raises a further question. If independent knowledge bases, operating in different contexts, produce the same pattern, then the explanation cannot lie in isolated methodological choices. It must lie in the way the knowledge base itself is constituted and transmitted. The concept of consensus therefore serves as a bridge. It marks the point at which individual observations coalesce into a structural finding, and it prepares the ground for the central problem addressed in this paper: not only that measurement is absent, but that its absence is consistently reproduced.

The remainder of the analysis builds on this foundation. Having established that the pattern is one of consensus rather than variation, the focus turns to its implications and, more importantly, to its persistence. The key issue is not simply that HTA fails to meet the conditions required for measurement, but that it does so in a way that is stable across domains and resistant to correction. Understanding this persistence requires moving beyond description to explanation.

## **THE AUSTRALIAN HTA RESEARCH LANDSCAPE**

The Australian health technology assessment (HTA) research landscape is characterized by a relatively small number of academic centers that play a central role in the development, application, and dissemination of HTA methods. These centers are typically embedded within universities, often located in schools of public health, health economics, or population health, and are closely aligned with national policy requirements. Their outputs include economic evaluations, decision-analytic models, methodological research, and contributions to reimbursement submissions. While differing in institutional structure and research emphasis, they collectively define the academic knowledge base from which HTA practice in Australia is drawn.

Five research center knowledge bases are evaluated. There are the Centre for Health Economics Research and Evaluation (CHERE) at the University of Technology Sydney, the Centre for Health Economics (CHE) at Monash University, the Melbourne School of Population and Global Health (MSPGH) at the University of Melbourne, the Health Economics group (HEG) at the University of Adelaide and the Economic and Research Modelling Unit (HERMU) at the University of Queensland.

These centers vary in their specific focus with some emphasizing applied economic evaluation and policy engagement, others concentrating on methodological development or population-level analysis but all operate within a common analytical framework centered on cost-effectiveness analysis and the use of preference-based outcome measures.

A defining feature of the Australian HTA landscape is its close integration with policy, particularly through the requirements of the Pharmaceutical Benefits Advisory Committee (PBAC). The PBAC mandates the use of cost-effectiveness analysis, typically expressed in terms of cost per QALY gained, as the basis for evaluating pharmaceutical submissions. This requirement exerts a strong influence on academic research. University-based HTA centers align their methods with these expectations, ensuring that their work is relevant to policy and capable of informing reimbursement decisions. As a result, there is a convergence between academic practice and policy frameworks, with the same constructs, utilities, QALYs, and incremental cost-effectiveness ratios serving as the core of both.

This alignment has important implications for the structure of the knowledge base. The use of standardized templates and methodological guidelines promotes consistency across institutions. Researchers are trained within programs that emphasize accepted HTA methods, and their work is evaluated through peer review processes that reinforce adherence to these methods. Over time, this creates a stable and internally coherent framework in which variation occurs at the level of application with different diseases, interventions, and data sources but not at the level of underlying constructs.

Despite this apparent diversity in application, the core methodological elements remain constant. Economic evaluations rely on preference-based measures of health-related quality of life, typically derived from instruments such as the EQ-5D, and these measures are combined with survival or time-based outcomes to produce QALYs. Decision-analytic models are used to extrapolate beyond observed data, incorporating assumptions about disease progression, treatment effects, and patient behavior. The outputs of these models are expressed in numerical terms and interpreted as evidence to support decision-making.

The knowledge base that emerges from this landscape is therefore both extensive and coherent. It encompasses a wide range of studies, methodological discussions, and policy-relevant analyses, all operating within a shared framework. This framework is not explicitly defined in terms of its measurement properties, but it is implicitly structured by the requirements of HTA practice. The interrogation of this knowledge base, using the canonical diagnostic, does not focus on the volume or diversity of output, but on the underlying assumptions that govern the use of numbers within this framework.

The inclusion of multiple Australian HTA research centers in this analysis serves to test whether these assumptions vary across domains. Given the differences in institutional context, research focus, and academic orientation, one might expect some variation in how measurement principles are recognized or applied. However, as the results demonstrate, this variation does not occur at the level of measurement. The knowledge base, while diverse in its applications, exhibits a consistent structure in its treatment of the conditions required for measurement.

This consistency is central to the argument that follows. The Australian HTA research landscape provides a suitable and robust context for examining whether the absence of measurement is an isolated feature or a systematic property. By selecting knowledge bases that differ in form but share a common methodological framework, the analysis is able to distinguish between local variation and structural consistency. The findings indicate that it is the latter that prevails.

## **INTERROGATION FRAMEWORK AND ANALYTICAL METHOD**

The analysis presented in this paper is based on a standardized interrogation framework designed to evaluate whether a knowledge base recognizes the conditions required for valid measurement. The framework is grounded in the axioms of representational measurement and operationalized through a fixed diagnostic instrument comprising 24 canonical statements. Each statement is constructed such that it is unequivocally true or false under these axioms. The purpose is not to solicit opinion or interpretation, but to determine whether the knowledge base, as represented in its corpus of outputs and methodological conventions, reinforces or rejects propositions that are fundamental to measurement.

Each Australian HTA research center is treated as a defined knowledge domain. This domain includes its published research, methodological approaches, teaching materials, and the implicit assumptions that can be inferred from these sources. The interrogation is conducted by presenting the canonical statements sequentially, with the knowledge base instructed to evaluate each statement in the context of its own established practices. The constraint is important: the knowledge base is not invited to reinterpret the statement or introduce external qualifications, but to assess it as it would be understood within the domain. This ensures that the response reflects the internal structure of the knowledge base rather than the flexibility of language.

Responses are expressed as categorical probabilities drawn from a fixed and deliberately limited set. These categories 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.55, 0.60, 0.65, 0.75, 0.80, 0.85, 0.90, and 0.95 are selected to avoid spurious precision while allowing meaningful differentiation in the strength of endorsement. The assigned probability represents the extent to which the knowledge base supports the proposition embodied in the statement. Low probabilities indicate weak endorsement or effective rejection; high probabilities indicate strong endorsement. Intermediate values capture limited or partial recognition.

These categorical probabilities are then transformed into normalized logits using the standard log-odds transformation. The logit, defined as the natural logarithm of the odds ratio, maps bounded probability values onto a continuous scale. For the purposes of this analysis, logits are normalized to a fixed range ( $\pm 2.50$ ) using established rounding conventions. This normalization ensures comparability across interrogations and prevents extreme values from distorting interpretation. Equal differences in logits correspond to equal differences in the strength of endorsement, allowing the identification of patterns that are not apparent from ordinal probability categories alone.

Within a given interrogation, the set of probability–logit pairs across the 24 statements constitutes a profile of the knowledge base. Differences in these values indicate the relative strength with which individual propositions are supported or rejected. More importantly, the distribution of

logits across true and false statements reveals whether the knowledge base aligns with or diverges from the axioms of measurement. A profile in which true statements are strongly endorsed and false statements rejected would indicate alignment with measurement principles. The inverse pattern indicates divergence.

Across interrogations, the use of a fixed diagnostic instrument, constrained probability categories, and a consistent logit transformation ensures that comparisons are valid. Because each knowledge base is evaluated under identical conditions, differences in results can be attributed to differences in the knowledge base rather than variation in method. When multiple knowledge bases produce similar profiles, the interpretation is that they share an underlying structure.

It is important to emphasize that the objective of this framework is not to produce a summary score, ranking, or index. The focus is on structure, not aggregation. Each statement is evaluated independently, and the resulting probabilities and logits are interpreted in relation to the truth value of the statement under representational measurement. The absence of a composite score avoids the introduction of additional assumptions regarding weighting or dimensionality, which would themselves require justification.

The framework also incorporates an implicit test of falsifiability. Statements concerning the rejection of non-falsifiable claims and the role of models in generating testable propositions are included alongside measurement axioms. This allows the interrogation to assess not only whether measurement is recognized, but whether the knowledge base acknowledges the conditions required for scientific testing. The relationship between these elements is central to the interpretation of results.

The application of this framework to Australian HTA research centers follows the same protocol as that used in the Canadian analysis. This ensures that the results are directly comparable across jurisdictions. Any observed similarities or differences can therefore be interpreted as properties of the knowledge base rather than artifacts of the method. The consistency of the approach is essential for identifying whether the patterns observed are local or structural.

In summary, the interrogation framework provides a systematic and reproducible means of evaluating the alignment of HTA knowledge bases with the axioms of representational measurement. By combining categorical probabilities with normalized logits, it allows for both qualitative and quantitative assessment of endorsement patterns. When applied across multiple domains, it reveals whether those patterns are variable or consistent, and thereby supports the transition from individual analysis to structural interpretation.

## **RESULTS: A CONSENSUS PATTERN OF FALSE MEASUREMENT INVERSION**

The results of the five Australian HTA academic knowledge base interrogations are summarized in Table 1. For each of the 24 canonical statements, the table identifies whether the statement is true or false under the axioms of representational measurement, together with the observed range of categorical probabilities and the corresponding range of normalized logits across the five domains. As in the Canadian and US analyses, the purpose of presenting ranges is to make explicit

both the level and the consistency of endorsement. The question is not whether individual values differ, but whether the pattern of endorsement of representational measurement varies in direction or structure across knowledge bases.

**TABLE 1: SUMMARY OF STATEMENT RESPONSES: FIVE AUSTRALIAN RESEARCH CENTRE KNOWLEDGE BASE DOMAINS**

<b>STATEMENT</b>	<b>RESPONSE 1=TRUE 0=FALSE</b>	<b>ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY RANGE</b>	<b>NORMALIZED LOGIT (IN RANGE +/- 2.50) RANGE</b>
INTERVAL MEASURES LACK A TRUE ZERO	1	0.20 to 0.25	-1.40 to -2.25
MEASURES MUST BE UNIDIMENSIONAL	1	0/15 to 0.20	-1.60 to -1.40
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.20
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.80 to 0.85	+1.40 to +1.75
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.85 to 0.90	+1.75 TO +2.20
EQ-5D-3L PREFERENCE ALGORITHMS CREATE INTERVAL MEASURES	0	0.80 to 0.85	+1.40 to +1.75
THE QALY IS A RATIO MEASURE	0	0.85 to 0.90	+1.75 to +2.20
TIME IS A RATIO MEASURE	1	0.90 to 0.95	+2.20 to +2.50
MEASUREMENT PRECEDES ARITHMETIC	1	0.10	-2.20
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.85	+1.75
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.10	-2.20
THERE ARE ONLY TWO CLASSES OF MEASUREMENT LINEAR RATIO AND RASCH LOGIT RATIO	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50
SUMMATION OF LIKERT QUESTION SCORES CREATES A RATIO MEASURE	0	0.85	+1.75
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.80 to 0.85	+1.40 to +1.75

CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.15 to 0.20	-1.60 to -1.40
QALYS CAN BE AGGREGATED	0	0.85 to 0.90	+1.75 to +2.20
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.55 to 0.65	+0.50 to +0.85
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.80 to, 0.85	+1.40 to +1.75
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.55 to 0.60	+0.50 to +0.70
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FOR ASSESSING THERAPY IMPACT FOR LATENT TRAITS	1	0.05	-2.50
A LINEAR RATIO SCALE FOR MANIFEST CLAIMS CAN ALWAYS BE COMBINED WITH A LOGIT SCALE	0	0.35	-1.25
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.25	-1.90
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

Note: Where no range is reported all five knowledge bases were allocated to the same categorical endorsement probability.

The summary profile for the five Australian HTA research centers presents a remarkably consistent and highly structured pattern of endorsement that leaves little doubt as to the nature of their shared knowledge base. What emerges is not random variation across institutions, but a tightly clustered set of beliefs that reflect a common intellectual inheritance; one grounded in the reference case framework and its associated modeling practices. The ranges of probabilities and logits are narrow across almost all statements, indicating that these centers are not merely similar; they are effectively interchangeable in their adherence to the same conceptual system of measurement inversion.

At the core of this system is a systematic under-endorsement of the foundational axioms of representational measurement. Statements that should be trivially accepted within any scientific measurement framework such as the requirement that measurement precedes arithmetic or that multiplication requires a ratio scale are assigned probabilities of 0.10, corresponding to logits of -2.20. This is not a marginal hesitation; it is a near-total absence of endorsement. Similarly, the proposition that there are only two permissible classes of measurement such as linear ratio for manifest attributes and Rasch logit ratio for latent constructs are assigned the lowest possible probability (0.05; -2.50 logits). This indicates that the conceptual foundation required for lawful

measurement is almost entirely missing from the knowledge base. There is no awareness of the axioms of representational measurement

In contrast, statements that are demonstrably false under representational measurement theory are strongly endorsed. The belief that EQ-5D-3L preference algorithms create interval measures, that QALYs behave as ratio measures, and that summation of ordinal responses can generate ratio properties are all supported with probabilities in the range of 0.80 to 0.90, corresponding to logits between +1.40 and +2.20. This reversal where false propositions are strongly endorsed and true propositions rejected is the defining feature of what is termed measurement inversion.

The QALY occupies a central position in this inversion. Across all five centers, the rejection of the statement “the QALY is a ratio measure” is accompanied by high endorsement probabilities for its implicit opposite. This allows the QALY to function as the primary outcome metric in cost-effectiveness analysis, despite its failure to meet even the most basic requirements of dimensional homogeneity. The multiplication of time (a ratio scale) by utility scores (ordinal or, at best, unverified interval approximations) produces a quantity that cannot support arithmetic operations. Yet this quantity is treated as if it were a valid measure, and is further combined with cost aggregates to generate ICERs. The table shows that this belief is not contested across centers; it is uniformly reinforced.

Equally revealing is the complete rejection of Rasch measurement. Statements asserting the necessity of Rasch transformation for latent constructs, or the equivalence of Rasch rules with the axioms of representational measurement, are assigned probabilities of 0.05 across all centers. This indicates not merely neglect, but exclusion. The Rasch model, which would impose strict requirements for unidimensionality, invariance, and interval scaling, is incompatible with the multiattribute instruments and composite indices that dominate HTA practice. Its absence is therefore structurally necessary to maintain the current framework.

There is, however, a limited and somewhat inconsistent endorsement of general scientific norms. The statement that non-falsifiable claims should be rejected receives moderate support (0.55–0.65; logits +0.52 to +0.85), suggesting that these centers recognize, at least in principle, the importance of falsifiability. Similarly, there is modest agreement that the logit represents the natural logarithm of the odds ratio. Yet these acknowledgments do not translate into practice. Simulation models, which dominate HTA outputs, are treated as if they generate falsifiable claims, despite being inherently dependent on assumptions that cannot be empirically verified in a meaningful sense. The table captures this contradiction: strong endorsement of falsifiability as a principle, coupled with strong endorsement of modeling practices that violate it.

The narrowness of the probability ranges is itself significant. Across most statements, variation is limited to a span of 0.05 or 0.10, with corresponding tight logit intervals. This indicates that the five centers share a highly stable and homogeneous knowledge base. There is no evidence of dissenting positions or emerging alternatives within this group. The consistency suggests that the knowledge base is not being actively interrogated or revised, but rather reproduced through teaching, publication, and collaboration.

This has important implications for the evolution of objective knowledge. A scientific field advances through the formulation and testing of falsifiable claims, the correction of errors, and the refinement of measurement. The profile presented here suggests that none of these processes are operating effectively within the HTA research environment represented by these centers. Instead, we see a closed system in which core assumptions are insulated from challenge, and where numerical outputs are accepted as evidence without satisfying the conditions required for measurement.

In practical terms, this means that the outputs generated by these centers such as cost-effectiveness ratios, QALY estimates, and model projections cannot be considered credible in the scientific sense you have defined. They are internally consistent within the reference case framework, but they do not meet the standards of unidimensionality, invariance, or dimensional homogeneity. As such, they cannot support evaluable or replicable claims.

The conclusion is unavoidable. The five Australian HTA research centers examined here are not independent contributors to a diverse scientific field; they are components of a single, highly coherent memplex that has inverted the principles of measurement. Until that inversion is addressed—through the adoption of Rasch measurement for latent constructs and strict adherence to ratio scaling for manifest attributes—the field will remain locked in a cycle of numerical storytelling, unable to generate the kind of evidence required for genuine scientific progress.

## **REPRODUCTION OF A CLOSED KNOWLEDGE BASE**

The consistency of results across Australian HTA research centers requires explanation. As in the Canadian case, the interrogation reveals a stable pattern in which statements grounded in the axioms of measurement are weakly endorsed, while incompatible propositions are strongly reinforced. The key question is not whether this pattern exists, but why it persists across institutions that differ in size, location, and research focus. The answer does not lie in independent analytical judgment. It lies in the reproduction of a closed knowledge base.

The first element of this reproduction is entry through training. Researchers entering the HTA field do not begin with first principles of measurement. They are introduced to a set of accepted methods including cost-effectiveness analysis, QALYs, and decision modeling presented as established tools rather than as propositions requiring validation. Textbooks, graduate programs, and supervisory relationships transmit these methods as the foundation of the discipline. The question of whether the underlying constructs meet the requirements of measurement is not raised. As a result, new entrants adopt the framework as given. The conditions required for measurement are excluded at the point of entry.

The second element is methodological lock-in. Within HTA, certain constructs are not optional; they are required. The QALY, utility measurement, and the reference case define what counts as acceptable analysis. Research proposals, grant applications, and submissions to policy agencies are expected to conform to these templates. Deviation is not encouraged. This creates a closed methodological environment in which the same constructs are repeatedly applied. The interrogation results reflect this: strong endorsement of QALYs and cost-effectiveness claims is not the result of evaluation, but of compliance with established expectations.

The third element is reinforcement through publication. Academic journals and peer review processes play a central role in maintaining the framework. Studies that employ accepted HTA constructs are recognized as methodologically sound, while those that challenge these constructs face barriers to publication. The review process does not interrogate the measurement properties of utilities or QALYs; it assesses adherence to established methods. This creates a feedback loop. Researchers produce work that conforms to expectations, journals publish it, and the published literature reinforces the legitimacy of the framework. Over time, the framework becomes self-validating.

The fourth element is alignment with policy. In Australia, as in other jurisdictions, HTA research is closely linked to decision-making bodies such as the Pharmaceutical Benefits Advisory Committee (PBAC). Policy requirements specify the use of cost-effectiveness analysis and QALYs as the basis for evaluation; there is no interest in representational measurement. Academic research aligns with these requirements to ensure its continued relevance to and impact in Australian HTA. The result is a convergence between academic practice and policy expectations. The same constructs are used in both domains, reinforcing their status as the standard approach. The interrogation results show that this alignment does not extend to measurement principles. The constructs are accepted, but the conditions required to justify them are not examined.

A fifth element is the absence of falsification. In a scientific framework, claims are subject to empirical test. Measurement provides the basis for this process, allowing hypotheses to be confirmed or refuted. In HTA, the core constructs are not falsified. QALY-based claims and cost-effectiveness ratios are not tested against measurable attributes; they are outputs of models that incorporate assumptions about preferences, utilities, and disease progression. These assumptions are not subjected to measurement validation. As a result, the framework does not correct itself. It evolves through elaboration rather than through refutation.

These elements together define a system that reproduces itself. Training introduces the framework, methodological requirements enforce it, publication reinforces it, policy aligns with it, and the absence of falsification prevents its correction. The result is a knowledge base that is stable and internally coherent, but closed to the principles that would allow it to engage with measurement. The interrogation results are therefore not surprising. They are the expected outcome of a system in which the axioms of measurement are excluded from the outset.

It is important to emphasize that this reproduction does not require coordination or intent. Individual researchers operate within the framework as it is presented to them. The consistency of the results across Australian centers is not evidence of agreement in the sense of deliberate consensus. It is evidence that the same framework is being applied across domains. What appears as methodological alignment is, in fact, the replication of a shared structure.

The persistence of this pattern has implications beyond the Australian context. When the same results are observed across jurisdictions with different institutional arrangements, it suggests that the underlying framework is not local but global. The reproduction of measurement inversion is not confined to a particular country or set of institutions. It is a feature of HTA as a field; a global memplex of false measurement.

## **CONCLUSION**

The evidence assembled across the five Australian research centers points to a conclusion that is both unavoidable and unsettling: the global HTA memplex has reached the end of the road. This is not a matter of incremental methodological refinement or the need for improved data inputs. It is a structural failure rooted in the systematic rejection or more precisely, inversion of the axioms of representational measurement. The consistency of the probability and logit profiles leaves no room for ambiguity. Across institutions, there is a shared and deeply embedded commitment to practices that cannot support measurement, cannot support arithmetic, and therefore cannot support scientifically credible claims.

What has been revealed is not simply a collection of errors, but a closed intellectual system. Within this system, composite constructs such as utilities and QALYs are treated as if they were measurable quantities, simulation models are treated as if they generate empirical evidence, and cost-effectiveness ratios are treated as if they have interpretive meaning. Each of these elements depends on assumptions that violate fundamental measurement principles, yet each is reinforced through training, publication, and policy engagement. The result is a self-sustaining framework in which numbers confer authority, regardless of whether they correspond to anything that can be measured.

The implications are profound. A discipline that does not respect the conditions for measurement cannot claim to produce evidence. A framework that cannot generate falsifiable, replicable claims cannot contribute to the evolution of objective knowledge. The continued reliance on QALYs, ICERs, and reference case simulations is therefore not merely misguided; it is indefensible. The appeal to tradition, consensus, or policy relevance cannot rescue a framework that fails at the most basic level of scientific legitimacy.

This is the point at which continuation becomes untenable. The global HTA memplex cannot be repaired from within because its core assumptions are the source of the problem. What is required is not reform, but replacement: a reconstruction of HTA grounded in lawful measurement, single-attribute value claims, and protocols that support empirical evaluation and replication. The alternative is to persist with a system that produces the appearance of evidence while systematically denying the possibility of 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.

## **WEBSITE REFERENCES**

**AUSTRALIA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION (CHERE) UNIVERSITY OF TECHNOLOGY SYDNEY**

**AUSTRALIA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE CENTRE FOR HEALTH ECONOMICS (CHE) MONASH UNIVERSITY**

**AUSTRALIA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE HEALTH ECONOMICS GROUP (HEG) UNIVERSITY OF ADELAIDE**

**AUSTRALIA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE MELBOURNE SCHOOL OF POPULATION AND GLOBAL HEALTH (MSPGH) UNIVERSITY OF MELBOURNE**

**AUSTRALIA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE HEALTH ECONOMICS AND RESEARCH MODELLING UNIT (HERMU) UNIVERSITY OF QUEENSLAND**