

Convergence

RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**UNITED KINGDOM: CONVERGENCE ON FALSE
MEASUREMENT - THE *JOURNAL OF HEALTH
ECONOMICS* AND THE *INTERNATIONAL JOURNAL OF
TECHNOLOGY ASSESSMENT IN HEALTH CARE***

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

LOGIT WORKING PAPER No 79 FEBRUARY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper compares the probability–logit profiles of two apex journals in health technology assessment and outcomes research: the *Journal of Medical Economics* and the *International Journal of Technology Assessment in Health Care*. Using a 24-item canonical diagnostic grounded in representational measurement theory, the analysis evaluates whether foundational scale-type axioms operate as binding constraints within their respective knowledge bases. The results demonstrate structural convergence. In both journals, propositions asserting unidimensionality, the precedence of measurement over arithmetic, the necessity of ratio scales for multiplication, and the equivalence of Rasch rules to representational measurement axioms collapse into the negative or floor region. In contrast, statements endorsing QALY aggregation, composite utility arithmetic, and the dimensional homogeneity of cost-per-QALY ratios cluster strongly in the positive region. This asymmetry indicates that arithmetic operations are reinforced while the axioms that make such operations meaningful are not. The comparison therefore reveals not isolated editorial weakness but systemic exclusion of measurement discipline as a governing principle in apex HTA publishing. The findings raise fundamental questions about falsifiability, replication, and the duty of care owed to clinicians and patients when quantitative claims inform therapeutic and reimbursement decisions.

INTRODUCTION

A comparison between apex journals such as the *Journal of Medical Economics* and the *International Journal of Technology Assessment in Health Care* matters because it tests whether the observed measurement failures are incidental or systemic. A single journal might be criticized as editorially idiosyncratic, theoretically narrow, or uneven in methodological rigor. But when independent leading journals exhibit the same probability–logit asymmetries with negative reinforcement of representational measurement axioms and positive reinforcement of composite utility arithmetic, the diagnosis shifts from local weakness to structural convergence. The issue is no longer what one editorial board believes; it is what the HTA knowledge base collectively endorses as legitimate quantification. A side-by-side logit comparison therefore serves as an empirical probe of the memplex itself. It reveals whether arithmetic without measurement is an anomaly or a governing rule across the evaluative ecosystem.

1. STRUCTURAL SYMMETRY: CONVERGENT LOGIT ARCHITECTURE

When the probability–logit profiles of the *Journal of Medical Economics* and the *International Journal of Technology Assessment in Health Care* are compared, the most important finding is structural symmetry. The two journals do not differ in kind; they align in pattern. The same propositions collapse into the negative region. The same propositions are reinforced in the positive region. The same asymmetry appears between foundational measurement axioms and utility-based arithmetic. This symmetry is more significant than the precise magnitude of individual logits.

In both journals, propositions that should function as boundary conditions—measurement precedes arithmetic, multiplication requires a ratio scale, measures must be unidimensional—do

not operate as structural constraints. Their endorsement probabilities translate into negative logits, indicating that they are not binding principles within the knowledge base. By contrast, propositions that license cost-per-QALY arithmetic, aggregation of QALYs, dimensional homogeneity of the QALY, acceptance of composite scoring as measurement are positively reinforced.

The implication is not that both journals are confused about measurement. The implication is that both operate within the same evaluative architecture. That architecture permits arithmetic to proceed independently of scale-type verification. When two apex journals converge in this pattern, the issue cannot be dismissed as editorial preference or disciplinary nuance. It indicates epistemic consolidation around a shared framework. The logit profiles are not isolated assessments; they are structural diagnostics. Together they demonstrate that contemporary HTA publishing is internally coherent but externally incoherent relative to representational measurement theory. The symmetry itself is the finding.

2. RASCH COLLAPSE PATTERNS: EFFECTIVE NON-POSSESSION

The most decisive common feature in both journals is the collapse of Rasch-related propositions. Statements asserting that lawful transformation of ordinal responses requires Rasch rules, that the Rasch logit ratio scale is the only basis for measuring latent traits, and that Rasch rules are equivalent to the axioms of representational measurement fall to floor or near-floor values. In this diagnostic framework, -2.50 denotes effective non-possession. The proposition does not operate as a governing constraint within the evaluated corpus.

This matters because Rasch measurement is not a stylistic option within psychometrics. It is the only transformation model that produces invariant, interval-level measures from ordered categorical data under strict conditions. Without Rasch constraints, summation of item scores remains ordinal. Weighting does not create interval properties. Preference aggregation does not produce ratio measures. When both journals exhibit floor collapse for Rasch propositions, they are not selecting an alternative lawful transformation model. They are operating without one.

The consequence is that latent constructs are treated as measurable quantities without demonstration of invariance or unidimensionality. The journals publish analyses in which composite indices function as if they were measures, yet the structural requirement for lawful transformation is absent. The collapse pattern therefore identifies not methodological variation but foundational omission. Rasch principles are not contested; they are excluded. In both journals, this exclusion is systematic, not incidental.

3. UTILITY ENDORSEMENT CLUSTERS: REINFORCED COMPOSITE ARITHMETIC

In contrast to the Rasch collapse, both journals display strong positive reinforcement for QALY-centered propositions. Statements that QALYs can be aggregated, that EQ-5D algorithms generate interval measures, that QALYs are ratio measures, and that summated subjective responses produce ratio scales cluster in the high positive region. These are not weak endorsements. They demonstrate active reinforcement of composite utility arithmetic.

The symmetry is telling. Where Rasch transformation collapses to -2.50 , QALY aggregation rises to $+2.20$. Where unidimensionality is weakly reinforced, dimensional homogeneity of the QALY is affirmed. This inversion defines the epistemic structure of the memplex. The constructs that require scale discipline are neglected; the constructs that permit cost-effectiveness arithmetic are legitimized.

Importantly, this reinforcement does not prove the QALY is a ratio measure; it shows that the journals treat it as such. The logit instrument measures endorsement within the knowledge base, not ontological truth. High positive logits indicate that composite utility arithmetic functions as an accepted rule. The asymmetry therefore demonstrates belief consolidation around multiattribute constructs. It is not simply that measurement axioms are underemphasized; it is that composite arithmetic is positively institutionalized. Both journals participate in this reinforcement, and that participation defines their quantitative posture.

4. FALSIFICATION AND SIMULATION: CATEGORY ERROR AT SCALE

Both journals affirm in principle that non-falsifiable claims should be rejected. Yet both endorse, implicitly or explicitly, the legitimacy of reference case simulations as evidentiary outputs. This duality reveals a shared category error. Simulation outputs are recalculable under alternative parameter values; they are not empirically refutable claims about the world. Sensitivity analysis varies assumptions within a closed model; it does not expose a claim to possible empirical disconfirmation.

When simulation-generated cost-per-QALY ratios are treated as evidence, falsification becomes procedural rather than empirical. Recalibration replaces rejection. Adjustment replaces refutation. The journals' logit profiles show moderate reinforcement of falsification as an abstract principle but strong reinforcement of constructs that cannot be falsified in the Popperian sense. This coexistence is structurally inconsistent.

The implication is that both journals operate with a softened conception of scientific testing. Claims are evaluated for internal consistency, parameter robustness, or methodological conformity, not for empirical vulnerability. This matters because the evolution of objective knowledge depends on exposure to error. If the quantitative constructs themselves lack lawful measurement properties and the projections derived from them cannot be empirically falsified, then the evolutionary mechanism is compromised at two levels: scale and testing. The symmetry between the journals confirms that this compromise is not isolated.

5. DUTY OF CARE AND ECOSYSTEM DIAGNOSIS

The convergence of these patterns across apex journals shifts the argument from critique of editorial practice to diagnosis of an evaluative ecosystem. Journals shape what counts as legitimate evidence. They influence training, guideline development, and policy adoption. When leading journals converge in neglecting representational measurement axioms while reinforcing composite utility arithmetic, the downstream implications are substantial.

Duty of care in health system decision-making requires that therapy impact claims be grounded in unidimensional attributes, expressed on admissible scales, and exposed to empirical testing. If journals institutionalize constructs that do not satisfy these conditions, they normalize quantitative practices that lack measurement discipline. Physicians and patients are then served by projections and composite indices rather than lawful measures.

The comparison of JME and IJTAHC shows not divergence but alignment. Both participate in a stabilized quantitative belief system in which arithmetic precedes measurement and simulation substitutes for empirical exposure. This is not a matter of tone or rhetorical excess; it is demonstrated by homologous logit patterns. The journals are not failing independently. They are succeeding coherently within a shared framework that excludes representational measurement as a binding constraint. That convergence defines the contemporary HTA memplex.

CONCLUSION

The side-by-side logit profiles of JME and IJTAHC demonstrate epistemic consolidation rather than diversity. Both journals reinforce composite utility arithmetic while neglecting the axioms that determine when arithmetic is admissible. Rasch transformation statements collapse to floor values, unidimensionality is weakly reinforced, and cost-per-QALY constructs are treated as if they were ratio measures. The pattern is symmetrical and systemic. It cannot be attributed to isolated misunderstanding or editorial accident. Instead, it reveals a stabilized evaluative framework in which arithmetic conventions precede measurement discipline.

The consequence is not merely theoretical. When apex journals institutionalize quantitative claims that lack lawful scale properties and are insulated from empirical falsification, the evolution of objective knowledge is impaired. Replication becomes repetition within a closed architecture rather than independent testing against empirical structure. In such a setting, duty of care is compromised: clinicians and patients are served by projections and composite indices rather than measurement-valid claims. The comparison therefore does more than critique journals; it exposes the structural foundations of the contemporary HTA memplex and the need to re-anchor quantitative evaluation in representational measurement.

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.