MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

FOUNDATIONS OF SCIENTIFIC HEALTH TECHNOLOGY ASSESSMENT

INDICATIVE COMPETENCY QUESTIONS

Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 867 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper proposes a new framework for defining scientific competence in health technology assessment (HTA). Rather than assessing familiarity with utilities, quality-adjusted life years (QALYs), cost-effectiveness analysis, and reference-case simulation models, it identifies the minimum competencies required for the scientific evaluation of therapeutic interventions. The framework is motivated by a series of more than 230 large language model interrogations of HTA knowledge bases associated with research centers, academic programs, government agencies, journals, and professional organizations. These investigations consistently identified measurement inversion, where arithmetic precedes measurement, and curriculum inversion, where students are trained in analytical techniques without first acquiring the principles that determine whether those techniques can support valid quantitative claims.

The paper argues that scientific HTA must be grounded in representational measurement, the theory of measurement scales, lawful ratio measurement, admissible arithmetic, the distinction between manifest and latent attributes, Rasch measurement, and the formulation of evaluable, replicable, and falsifiable therapy impact claims. It further argues that the historical reliance on assumption-driven reference-case simulation models cannot compensate for the absence of lawful measurement or replace the scientific method. Simulation may explore assumptions, but it cannot create measures or generate scientific evidence from quantities that fail the requirements of representational measurement.

To operationalize these principles, the paper presents twenty-five cumulative competency questions that define the minimum knowledge expected of faculty members and graduate students claiming competence in scientific HTA. The questions progress from foundational concepts in measurement science to protocol-driven, falsifiable assessments of therapy impact. They are intended both as an assessment of individual competence and as an audit of existing HTA curricula. A curriculum capable of preparing students to answer these questions demonstrates adherence to the standards of normal science; a curriculum that cannot do so provides evidence of curriculum inversion. The paper concludes that these competencies provide the educational foundation for replacing the traditional reference-case paradigm with a measurement-based framework for HTA.

INTRODUCTION

Health technology assessment occupies a unique position within health economics. For more than four decades it has developed and promoted a framework for evaluating therapeutic interventions that has largely excluded the principles of representational measurement and the theory of measurement scales from its analytical foundations. Rather than establishing whether the quantities used to assess therapy impact satisfy the requirements for lawful measurement before undertaking arithmetic, HTA has institutionalized measurement inversion, where arithmetic precedes measurement, and curriculum inversion, where students are taught utilities, QALYs, cost-effectiveness analysis, and reference-case simulation models before they are introduced to the scientific principles that determine whether those methods can support valid quantitative claims.

The consequence is unique among the quantitative sciences: an entire discipline whose central analytical framework rests upon numerical constructions that were never first demonstrated to satisfy the standards required for scientific measurement.

The purpose of this assessment, a set of 25 Indicative Competency Questions, is to provide a practical means of assessing whether a contemporary health technology assessment (HTA) curriculum equips students and faculty with the knowledge required for measurement-based science. They are not intended as an examination of specialist expertise or advanced mathematical ability. Rather, they address the foundational concepts that must be understood before any quantitative claim regarding therapeutic impact can be regarded as scientifically credible. A graduate who has completed a modern HTA program should be able to answer these questions confidently if the curriculum genuinely prepares students to evaluate the scientific legitimacy of the analytical methods they are taught..

The need for such an assessment follows directly from the findings of the measurement inversion and curriculum inversion interrogations reported in over 230 Logit Working Papers. Those interrogations demonstrated a strikingly consistent pattern across HTA research centers, academic programs, government agencies, journals, and professional organizations. Foundational concepts such as representational measurement, scale theory, ratio measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, manifest and latent attributes, and Rasch measurement were either absent or only weakly represented, while utilities, QALYs, cost-effectiveness ratios, and reference-case simulation models continued to receive strong endorsement. These findings raise an obvious question: does the current HTA curriculum provide students with the conceptual tools necessary to recognize these deficiencies? Or does it prepare them only for numerical storytelling?

The competency questions are designed to recognize these deficiencies. Each addresses an issue that should form part of the core intellectual framework of a measurement-based HTA curriculum. They begin with the nature of attributes and the properties of measurement scales before progressing to ratio measurement, representational measurement, manifest and latent attributes, Rasch measurement, and the conditions required for evaluable and falsifiable claims. Together, they represent the minimum body of knowledge expected of graduates who claim competence in the scientific evaluation of therapeutic value.

The questions also provide a straightforward audit of existing curricula. Universities, research centers, faculty members, graduate students, manufacturers, and HTA agencies may use them to assess whether their current educational programs meet the standards expected of a quantitative scientific discipline. A curriculum that prepares students to answer these questions demonstrates that measurement precedes arithmetic and that quantitative claims rest upon lawful measurement. A curriculum that cannot do so reveals important educational gaps and provides evidence of curriculum inversion. The questions therefore serve not only as an assessment of faculty and student competence but also as an assessment of the scientific credibility of the curriculum itself and its capacity to support the transition from the reference-case paradigm to a measurement-based framework for HTA.

FOUNDATIONS OF SCIENTIFIC HTA

For many practitioners, the term *scientific HTA* will appear almost self-evident. For more than four decades, health technology assessment has been presented as a scientifically rigorous framework for evaluating therapeutic value through increasingly sophisticated statistical analyses, economic models, and lifetime simulations. Methodological complexity has come to be equated with scientific credibility. Within this paradigm, the reference-case simulation model has become the defining characteristic of HTA, with its outputs routinely described as evidence to support reimbursement and policy decisions.

The historical rationale for the reference-case framework is well understood. At the time of product launch, evidence regarding long-term therapeutic impact is necessarily limited. Decision makers therefore sought a method that combined available clinical evidence with explicit assumptions regarding disease progression, treatment persistence, resource utilization, costs, and future outcomes. Reference-case simulation models emerged as a pragmatic solution, allowing projections of long-term costs and benefits before comprehensive empirical evidence became available.

This rationale, however, does not address the central issue identified in the present assessment. The criticism is not that simulation models employ assumptions. Every scientific model incorporates assumptions. The criticism is that the quantities entering the HTA model were never demonstrated to satisfy the requirements of measurement before being subjected to arithmetic operations [1]. Utility scores were assumed to possess ratio properties, QALYs were constructed without demonstrating dimensional homogeneity, and these quantities were then propagated through increasingly sophisticated simulation models. The standards of representational measurement were bypassed from the outset.

The issue is therefore not that the reference case relies upon assumptions. The issue is that those assumptions are applied to quantities whose measurement status has never been established. In every other quantitative science, the first question is whether the quantities entering a model are lawful measures. In HTA, that question was largely bypassed. Attention shifted immediately to modelling, uncertainty analysis, sensitivity analysis, and decision thresholds without first demonstrating that the underlying quantities satisfied the standards required for quantitative science.

From about 1850 to 1970, science progressively formalized measurement. Stevens, Rasch, Krantz, Luce, Suppes, and Tversky did not create an obstacle to quantitative inquiry; they established its foundations. At precisely the moment when those foundations had been codified, the HTA reference-case paradigm emerged. Rather than building upon this century of progress, it largely ignored it. Utilities, QALYs, and simulation models were introduced as though the preceding history of measurement science had never occurred. The result was not an extension of quantitative science but a reversal of it: arithmetic was elevated above measurement, numerical construction replaced lawful measurement, and assumption-driven modelling displaced falsifiable empirical claims. HTA therefore represents not the culmination of scientific progress but an interruption of it.

An equally important consequence was the abandonment of the scientific method itself. Rather than proposing explicit claims regarding therapy impact that could be evaluated, replicated, and potentially falsified in real populations, the reference-case paradigm generated hypothetical projections derived from assumed model structures and assumed parameter values. The objective was no longer to test claims against observation but to explore the implications of assumptions. Cost-effectiveness estimates became products of simulation rather than propositions capable of empirical refutation. In doing so, HTA departed not only from the standards of representational measurement but also from the Popperian requirement that scientific claims must be exposed to the possibility of failure [2,3] A framework that does not formulate claims capable of falsification cannot claim to generate scientific evidence, regardless of the apparent sophistication of its mathematics.

Consequently, the problem is not one of modelling but of measurement. A simulation cannot transform non-measures into measures. Equally, it cannot transform assumptions into scientific evidence. However sophisticated the computational framework, its outputs cannot possess greater scientific legitimacy than the quantities supplied as inputs or the assumptions upon which they depend. The failure therefore lies not in the existence of assumptions but in the neglect of the measurement standards and scientific method that should have governed construction of the model itself. Analytical sophistication became confused with scientific inquiry.

Science is not defined by the complexity of its mathematical techniques, the size of its datasets, or the sophistication of its computer models. It is defined by adherence to the principles governing the construction, evaluation, and testing of knowledge claims. Before any quantitative claim can be regarded as scientific, the quantities involved must satisfy the requirements of representational measurement. Before any model can generate evidence, its inputs must themselves be lawful measures. Before any conclusion can claim scientific authority, it must be capable of empirical evaluation, independent replication, and potential falsification.

The contemporary reference-case paradigm reverses this sequence. It begins with assumptions regarding utility scores, QALYs, and simulation models and then applies increasingly elaborate analytical methods to quantities whose measurement status has never been demonstrated. The resulting outputs are interpreted as evidence even though the conditions necessary for lawful quantitative claims remain unresolved. Sophisticated Markov modelling cannot compensate for the absence of measurement. Simulation manipulates assumptions; it does not create evidence.

The alternative proposed in this program restores HTA to the standards that govern every other quantitative scientific discipline. Every assessment begins by identifying the target attribute. The appropriate measurement framework is then established by distinguishing manifest from latent attributes and selecting the corresponding ratio measure. Only after measurement has been demonstrated can arithmetic be performed, hypotheses formulated, and therapy impact claims specified. Those claims must be prospectively defined in explicit protocols, subjected to empirical observation, capable of independent replication, and open to refutation if the evidence fails to support them. Science advances by testing claims against reality, not by refining assumptions within increasingly complex simulations.

Scientific HTA is therefore not defined by utilities, QALYs, cost-effectiveness ratios, or reference-case simulation models. It is defined by scales of measurement [4] ,representational measurement [5], lawful ratio measurement, admissible arithmetic, Rasch measurement [6] , prospectively specified and evaluable therapy impact claims, explicit protocols, empirical observation, independent replication, and falsifiability. These are the defining characteristics of normal science and the minimum standards expected of every quantitative discipline. Their absence from the traditional HTA paradigm explains not only the pervasive pattern of measurement inversion documented throughout these investigations but also the emergence of curriculum inversion and, ultimately, paradigm failure. The purpose of the reconstructed curriculum is to restore these scientific foundations to the center of HTA so that claims regarding therapy impact are grounded in lawful measurement rather than assumption, in empirical evidence rather than numerical construction, and in scientific inquiry rather than administrative convention.

THE CUMULATIVE NATURE OF THE MINIMUM COMPETENCY QUESTIONS

The 25 Indicative Competency Questions are presented in a deliberate sequence because the principles of measurement are cumulative. Each question introduces concepts that must be understood before the next can be answered. The sequence follows the logic of scientific inquiry itself. It begins with attributes, measurement, and the properties of measurement scales before progressing to representational measurement, ratio measurement, manifest and latent attributes, Rasch measurement, and the construction of evaluable and falsifiable claims. Each concept depends upon mastery of those that precede it.

The competency questions therefore do not ask students to evaluate the strengths and weaknesses of utilities or QALYs as though these were competing scientific alternatives. They are not. Once the foundational competencies have been mastered, these constructs cease to qualify as scientifically defensible measures and therefore have no place within a measurement-based curriculum. They become historical examples of what occurs when arithmetic is allowed to precede measurement rather than contemporary analytical tools.

There is an important parallel between these competency questions and the Rasch approach to instrument development. In Rasch measurement, questionnaire items are deliberately ordered along a continuum of increasing difficulty. Individuals demonstrate possession of the latent attribute by successfully responding to progressively more demanding items. Possession is therefore not inferred from a single response but from a consistent pattern of responses across an ordered hierarchy of item difficulty.

The same principle applies to the competency questions presented here. They are arranged in a logical sequence of increasing conceptual difficulty. Mastery of the early questions is a prerequisite for answering those that follow because each introduces concepts upon which the next depends. Possession of competency in measurement-based HTA is therefore demonstrated not by isolated factual knowledge but by the ability to progress through this hierarchy of increasingly demanding concepts. A respondent who can answer all 25 questions demonstrates possession of the conceptual framework required for measurement-based HTA. Conversely, failure to answer the early questions precludes meaningful engagement with the later ones. In this sense, the

competency questions provide not merely an examination of knowledge but an assessment of possession of the intellectual attributes required to practice HTA as a quantitative scientific discipline.

The 25 questions are:

Question 1: Why must measurement precede arithmetic?

Question 2: Explain why identifying the target attribute is the first step in constructing a scientifically valid assessment of therapy impact.

Question 3: Why is identifying the target attribute as manifest or latent the next requirement in constructing a scientifically valid therapy impact claim?

Question 4: Explain why therapy impact claims for manifest attributes require linear ratio measures, whereas therapy impact claims for latent attributes require Rasch logit ratio measures.

Question 5: Explain why the limitations on arithmetic imposed by the principal scales of measurement apply equally to measures of manifest and latent attributes.

Question 6: Explain how the properties of nominal, ordinal, interval, and ratio scales determine the arithmetic operations that may be performed on their measures.

Question 7: Why are multiplication and division scientifically admissible only for ratio measures?

Question 8: Explain why lawful multiplication and division require both ratio measurement and dimensional homogeneity?

Question 9: Why is the unidimensionality of the target attribute a necessary condition for representational measurement?

Question 10: Explain why the purpose of representational measurement is to ensure that quantitative claims correspond to measurable attributes rather than arbitrary numerical assignments.

Question 11: Why do health-state valuations fail the standards for measurement?

Question 12: Why are utility values no more than composite ordinal scores?

Question 13: Why does the QALY fail the requirement of dimensional homogeneity?

Question 14: Explain why Rasch measurement is the necessary and sufficient condition for constructing ratio measures of latent therapy impact from ordinal patient responses.

Question 15: The probabilistic structure of the Rasch model arises from the interaction between item difficulty and the respondent's level of possession of the latent attribute. Explain why this interaction is fundamental to Rasch measurement.

Question 16: Why is the fit of a provisional item pool to the Rasch model a necessary condition for constructing a ratio measure of a latent attribute? Explain

Question 17: Why does the Rasch estimate of latent attribute possession depend upon the respondent's complete pattern of questionnaire responses? Explain.

Question 18: Explain why the Rasch logit ratio scale provides the only scientifically defensible measure of therapy possession of a latent therapy attribute?

Question 19: Explain why the existence of these two distinct forms of ratio measurement, linear ratio and Rasch logit ratio, preclude the construction of a single composite measure of therapy impact?

Question 20. Why must every therapy impact claim be accompanied by a protocol that specifies the target attribute, measurement method, comparator, time horizon, and criteria for empirical evaluation? Explain

Question 21: Measurement is necessary, but science requires measurement plus hypothesis testing. Explain

Question 22: Why is falsifiability a necessary condition for scientific claims regarding therapy impact? Explain

Question 23: Why is a lawful ratio measurement a necessary but not sufficient condition for a scientifically credible therapy impact claim.

Question 24: Explain why reference-case simulation models cannot generate evaluable and falsifiable evidence regarding therapy impact.

Question 25: Explain why replacing the reference-case paradigm requires reconstruction rather than reform of HTA

INTERPRETING THE COMPETENCY QUESTIONS

The competency questions presented above are not intended to represent a comprehensive examination in health technology assessment. Their purpose is more fundamental. They define the minimum scientific knowledge required for participation in a measurement-based HTA curriculum. Each question addresses a concept that should be understood before students are introduced to the analytical techniques used to evaluate therapy impact. Together they provide an objective benchmark against which existing curricula, faculty expertise, research programs, and student competence may be assessed.

The significance of these questions extends beyond individual students. They provide an equally important audit of existing educational programs. A curriculum whose faculty cannot provide satisfactory answers to these questions cannot reasonably claim to prepare students for scientific HTA. More importantly, a curriculum that does not teach these concepts will inevitably continue to reproduce the measurement inversion and curriculum inversion documented throughout this series of interrogations. A scientific HTA transition program that does teach these concepts has been developed by Maimon Research LLC. Graduates of this program are equipped to respond to each of these 25 competency questions and make a productive contribution to the assessment of therapy impact [7].

The purpose of the competency assessment is therefore straightforward. It is to determine whether an HTA program has successfully transitioned from the traditional reference-case paradigm to a framework grounded in representational measurement and the standards of normal science. A curriculum that enables students to answer these questions has moved beyond utilities, QALYs, cost-effectiveness ratios, and reference-case simulation models. The competency questions are not simply an examination of knowledge; they are an indicator of whether the transition to scientific HTA has genuinely occurred. A student or faculty member unable to answer these questions cannot reasonably claim competence in scientific health technology assessment.

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] The End of the Reference Case: Reconstructing HTA. Logit Working Paper No 345 June 2026. https://maimonresearch.com/logit-working-paper-no-345-june-2026/

[2] Popper K. The Logic of Scientific Discovery. London: Hutchinson, 1959 (English translation of the 1934 German edition).

[3] Popper K. Objective Knowledge: An Evolutionary Approach. Oxford: Clarendon Press 1979.

[4] Stevens S. On the Theory of Scales of Measurement" *Science* 1946; 103 (2684) 677–680.

[5] Krantz D, Luce R, Patrick P, Tversky A. *Foundations of Measurement. Volume I: Additive and Polynomial Representations*. New York: Academic Press, 1971.

[6] Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed). New York: Routledge, 2021

[7] Maimon Research Transition Program https://maimonresearch.com/hta-reconstruction-program-and-fees/