

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



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**UNITED STATES: PARADIGM FAILURE AND
CURRICULUM INVERSION IN HEALTH
TECHNOLOGY ASSESSMENT**

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

LOGIT WORKING PAPER No 658 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

For more than four decades, health technology assessment (HTA) in the United States has been dominated by a reference-case paradigm built upon utility measurement, quality-adjusted life years (QALYs), cost-effectiveness ratios, and simulation modelling. While this framework is widely regarded as the foundation of modern therapeutic evaluation, its scientific legitimacy depends upon a prior question: whether the quantities entering these analyses satisfy the requirements of measurement. This study extends previous investigations of measurement inversion by examining the educational foundations of HTA through an assessment of curriculum inversion across seven leading United States HTA research centres.

The analysis employs a large language model interrogation framework to evaluate the curriculum knowledge bases associated with the Schaeffer Center, Harvard Center for Health Decision Science, CHOICE Institute, Duke-Margolis Institute, Leonard Davis Institute, Bloomberg School of Public Health, and Stanford Health Policy Center. Ten canonical curriculum statements derived from representational measurement theory are used to assess exposure to foundational concepts, including attributes, target attribute specification, scales of measurement, measurement status, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement, and falsifiability. Endorsement probabilities provide an estimate of the extent to which these concepts are represented within the educational environment.

The results reveal a strikingly consistent pattern of curriculum inversion across all seven centres. While concepts such as attributes and falsifiability receive strong endorsement, foundational principles of measurement science receive uniformly weak endorsement. Scales of measurement, representational measurement, unidimensionality, measurement status, latent attribute measurement, and the distinction between manifest and latent ratio measures occupy only a marginal position within the curriculum knowledge base. The findings suggest that students and researchers are routinely exposed to modelling, economic evaluation, utility assessment, and QALY construction without corresponding exposure to the measurement principles necessary to evaluate the scientific legitimacy of those activities.

The study argues that curriculum inversion provides the missing explanation for the measurement inversion previously identified in HTA research environments. Educational systems that neglect measurement theory inevitably reproduce research frameworks that assume rather than demonstrate measurement. Together, curriculum inversion and measurement inversion point to a broader condition of paradigm failure in which arithmetic displaces measurement and methodological convention displaces scientific validation. The paper concludes that the contemporary HTA paradigm cannot be sustained through incremental reform and requires reconstruction around lawful measurement. A transition is proposed toward a framework based on linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes, providing a foundation for evaluable, replicable, and falsifiable claims regarding therapeutic impact.

INTRODUCTION

The evidence accumulated to date leaves little room for ambiguity. Across a growing series of interrogations of HTA agencies, academic centers, professional organizations, journals, and educational programs, the dominant pattern has been one of measurement inversion. Rather than establishing measurement properties before undertaking arithmetic operations, the contemporary HTA framework proceeds in the opposite direction. Arithmetic is accepted as legitimate while measurement is assumed. Utility scores are treated as though they possess ratio properties. QALYs are constructed without demonstrating dimensional homogeneity. Simulation models manipulate quantities whose measurement status remains unresolved. The result is a framework built upon assumptions regarding measurement rather than measurement itself.

The persistence of this pattern raises an obvious question. How could a discipline sustain such assumptions for more than forty years without recognizing the implications of measurement theory? The most plausible explanation is curriculum inversion. If the concepts required to evaluate quantitative claims are absent from the educational environment, then the resulting research and policy frameworks will inevitably reproduce the same deficiencies. Measurement inversion in research practice is unlikely to be accidental. Rather, it reflects the intellectual foundations upon which researchers, analysts, and decision makers were trained.

Curriculum inversion occurs when students are taught how to apply analytical techniques before they are taught how to determine whether the quantities entering those analyses satisfy the requirements of measurement. Economic evaluation, utility assessment, QALY construction, cost-effectiveness analysis, and simulation modelling become the primary focus of instruction, while the principles that govern quantitative science receive little attention. Students learn how to manipulate numbers, construct models, and generate estimates, but are given limited exposure to representational measurement, scale theory, admissible arithmetic, dimensional homogeneity, unidimensionality, manifest and latent attributes, or Rasch measurement. The educational sequence mirrors the analytical sequence observed in measurement inversion: arithmetic precedes measurement.

This possibility is particularly important because research centers occupy a dual role within the HTA environment. They are not only producers of research but also centers of education and professional training. They shape doctoral programs, postdoctoral fellowships, professional workshops, methodological guidance, and the intellectual development of future researchers and policy analysts. If the concepts necessary to evaluate measurement claims are absent from these educational activities, then successive generations of HTA practitioners will inherit the same conceptual limitations as their predecessors.

The critical question therefore shifts from the research outputs of HTA institutions to the educational environments that sustain them. If measurement inversion characterizes the analytical framework, does curriculum inversion characterize the educational framework? Are students exposed to the principles of representational measurement before they encounter utilities, QALYs, and simulation models? Do curricula distinguish between ordinal, interval, and ratio scales? Are students taught the difference between manifest and latent attributes? Do they

understand why multiplication requires ratio measures and why latent attributes require Rasch measurement? Or are these concepts largely absent from the educational experience?

To address these questions, the present analysis turns from measurement inversion to curriculum inversion. Using the same large-language-model interrogation framework, the focus shifts from methodological propositions to curriculum exposure statements. The objective is straightforward: to determine whether the educational environments associated with leading United States HTA research centers provide students and researchers with the conceptual foundations necessary to evaluate quantitative claims. If these foundations are absent, then curriculum inversion provides the missing explanation for the persistence of measurement inversion. The former reproduces the latter. Together they reveal how a framework built upon arithmetic before measurement became institutionalized and sustained throughout the HTA enterprise.

RESEARCH CENTERS

The research centers that are the focus for this assessment of measurement inversion are the Schaeffer Center for Health Policy and Economics at the University of Southern California ¹, the Harvard Center for Health Decision Science ², the CHOICE Institute at the University of Washington ³, the Duke-Margolis Institute for Health Policy ⁴, the Leonard Davis Institute of Health Economics at the University of Pennsylvania ⁵, the Bloomberg School of Public Health at Johns Hopkins University ⁶ and the Stanford Health Policy Center ⁷.

If these centers demonstrate limited recognition of the standards governing curriculum, then the implications extend far beyond individual institutions. They raise fundamental questions regarding the scientific foundations of the reference-case paradigm itself. The objective of this assessment is therefore not merely to examine the curriculum knowledge profiles of these research groups, but to determine whether the intellectual foundations of US HTA remain consistent with the requirements of quantitative science.

STANDARDS FOR MEASUREMENT

The starting point for any scientific discipline that seeks to make quantitative claims is measurement. Before quantities can be manipulated mathematically, it must first be demonstrated that they possess the properties necessary to support the proposed arithmetic operations. This principle is fundamental to both the physical and social sciences. Measurement precedes arithmetic. Quantitative claims are valid only when the quantities involved satisfy the requirements of measurement. If these requirements are absent, arithmetic operations may still be performed, but the resulting outputs have no scientific standing as measures.

The importance of this principle is reflected in the theory of measurement scales ⁸. Not all numerical assignments possess the same properties. Nominal scales classify. Ordinal scales rank. Interval scales support differences between values. Ratio scales alone support the full range of arithmetic operations because they possess a true zero and permit proportional comparisons. Consequently, the admissibility of arithmetic depends upon scale type. Addition and subtraction require at least interval properties. Multiplication and division require ratio properties. This is not a matter of convention. It is a requirement imposed by the structure of measurement itself.

The central importance of ratio measurement follows directly from these considerations. Any claim involving multiplication, division, proportional comparison, growth rates, averages of ratios, or cost-effectiveness ratios requires quantities that possess ratio properties. If ratio measurement has not been demonstrated, these operations are inadmissible. Numerical manipulation cannot create measurement properties that are absent from the underlying scale. Arithmetic cannot substitute for measurement.

These requirements are formalized in the axioms of representational measurement⁹. Representational measurement provides the scientific framework that links empirical observations to numerical representations. Its purpose is to ensure that numerical assignments preserve the structure of the attribute being measured. Only when this correspondence is demonstrated can arithmetic operations be regarded as meaningful. The axioms of representational measurement therefore establish the conditions under which quantitative claims can be considered scientifically legitimate.

Among the most important of these requirements is unidimensionality. Measurement requires that an attribute represent a single dimension. If multiple attributes are combined into a composite score, numerical aggregation may be possible, but measurement has not necessarily occurred. Without unidimensionality there is no assurance that a numerical value represents a coherent quantity. The distinction between aggregation and measurement is therefore fundamental. Numbers can always be combined. Measures cannot be assumed.

Equally important is the distinction between manifest and latent attributes. Manifest attributes are directly observable and, where appropriately specified, support linear ratio measurement. Latent attributes are not directly observable and require a measurement model capable of estimating possession of the attribute. In the latter case, the required measure is the Rasch logit ratio scale¹⁰. These two forms of ratio measurement, linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes, provide the only scientifically defensible basis for quantitative claims regarding therapy impact.

Taken together, these principles establish a clear standard. Measurement must precede arithmetic. Scale properties determine admissible operations. Ratio measurement is required wherever proportional comparisons or multiplication are involved. Unidimensionality must be demonstrated before measurement can be claimed. Representational measurement provides the governing scientific framework. Any discipline seeking to generate quantitative claims must satisfy these requirements. Without them, numerical outputs remain constructions rather than measures, and quantitative claims become matters of assumption rather than science.

UNDERSTANDING CURRICULUM INVERSION

Curriculum inversion occurs when students are taught how to apply quantitative methods before they are taught the measurement principles that determine whether those methods are scientifically legitimate. Instead of first establishing the nature of attributes, the requirements of representational measurement, the properties of measurement scales, and the conditions governing admissible arithmetic, the sequence is reversed. Students are introduced to analytical

techniques and modelling procedures before they are exposed to the principles necessary to evaluate the validity of those techniques. In effect, measurement is assumed rather than taught.

This reversal lies at the heart of the contemporary HTA curriculum. In a scientifically coherent educational framework, measurement provides the foundation for quantitative reasoning. Students would first learn the distinction between nominal, ordinal, interval, and ratio scales. They would understand why measurement precedes arithmetic, why multiplication requires ratio properties, why unidimensionality is a prerequisite for measurement, and why manifest and latent attributes require different measurement frameworks. Only after these concepts were mastered would they be introduced to economic evaluation, utility assessment, cost-effectiveness analysis, and simulation modelling. Quantitative methods would follow measurement because their legitimacy derives from measurement.

Curriculum inversion abandons this sequence. Students are taught how to construct utilities, calculate QALYs, estimate cost-effectiveness ratios, and develop simulation models without first establishing whether the quantities entering these analyses satisfy the requirements of measurement. The burden of demonstrating measurement is displaced by the assumption that accepted methods are inherently legitimate. As a result, technical competence replaces conceptual understanding. Students learn how to perform calculations without learning how to determine whether those calculations are scientifically admissible.

The consequences are far-reaching. Once measurement is detached from education, the distinction between numbers and measures gradually disappears. Ordinal responses are treated as though they possess interval or ratio properties. Composite indices are interpreted as measures. Utility algorithms are accepted without questioning their measurement foundations. Mathematical sophistication becomes confused with scientific validity. Increasingly complex analytical techniques then conceal rather than illuminate the absence of measurement principles.

The problem is particularly evident in the treatment of utilities and QALYs. Students are taught how health-state valuations generate utility scores and how utility scores are combined with time to create QALYs. Yet they are rarely asked whether the utility component possesses the ratio properties required for multiplication. The educational emphasis falls upon calculation rather than measurement. The central question—whether the arithmetic is lawful—is largely bypassed.

The same pattern characterizes the teaching of simulation modelling. Students learn how to construct increasingly sophisticated models, undertake sensitivity analyses, and project outcomes across populations and time horizons. Yet simulation models do not create measurement. They merely manipulate quantities supplied to them. If those quantities lack measurement status, the model cannot remedy the defect. Without an understanding of measurement principles, students may come to regard methodological complexity as evidence of scientific rigor rather than recognizing that the validity of any model depends upon the validity of its inputs.

The significance of curriculum inversion therefore extends beyond individual courses or educational programs. It represents a reversal of the educational process through which scientific disciplines reproduce themselves. Rather than teaching the principles that govern quantitative claims and then introducing analytical methods, the curriculum begins with methods and leaves

the foundations largely unexplored. Students learn what to do before they learn why it is permissible to do it.

This is why curriculum inversion is so important. It provides a compelling explanation for the persistence of measurement inversion across HTA research, policy, and practice. Researchers can only challenge assumptions if they have been exposed to the concepts necessary to recognize them. If representational measurement, ratio measurement, dimensional homogeneity, manifest and latent attributes, and Rasch measurement are absent from the educational environment, then the assumptions embedded within utilities, QALYs, and reference-case simulation models are unlikely to be questioned. Curriculum inversion therefore becomes the mechanism through which measurement inversion is reproduced across successive generations of HTA practitioners.

More importantly, curriculum inversion identifies where reconstruction must begin. If HTA is to return to the standards of normal science, measurement must be restored to its proper place at the center of education. Until students are taught the principles that govern quantitative claims, the scientific legitimacy of HTA practice will remain fundamentally compromised.

.INTERROGATING THE KNOWLEDGE BASE

A large language model (LLM) is an artificial intelligence system trained on extensive collections of text to identify patterns, relationships, concepts, and structures within language. Rather than retrieving information from a fixed database, an LLM generates responses by estimating the most probable continuation of text based upon the knowledge patterns represented within its training corpus together with any additional information available at the time of interrogation. These additional sources may include publicly available web content, user-supplied documents, and accumulated contextual memory derived from prior interactions and analyses. Consequently, the effective knowledge base available to an interrogation extends beyond the model's original training cut-off and reflects the broader information environment to which the model has access.

In this analysis, “interrogation” refers not to discovering what an LLM believes. An LLM possesses neither beliefs nor opinions. Rather, interrogation refers to probing the content of a corpus-defined knowledge space. The objective is to characterize the concepts, assumptions, relationships, and analytical frameworks that are embedded within a defined body of literature and professional discourse. In the present study, the interrogation also draws upon approximately twelve months of accumulated HTA-related analyses undertaken using the same framework. This provides additional contextual information regarding the conceptual structure of the HTA literature and its treatment of measurement-related issues.

The corpus is defined before interrogation. Depending upon the objective of the study, it may consist of a journal, a national HTA agency, a professional organization, a methodological framework, a curriculum environment, a research Centre, or a collection of policy documents. Once the boundaries of the corpus have been established, the LLM is used to estimate its conceptual footprint. This approach allows identification of concepts that are strongly reinforced, weakly represented, neglected, misunderstood, or systematically excluded.

In the present assessment, the objective is precise: to determine the extent to which a given HTA knowledge base recognizes and reinforces the foundational principles of representational measurement theory. The core proposition under investigation is that measurement precedes arithmetic. No quantity should be treated as a number or subjected to mathematical operations unless the requirements of measurement have first been satisfied. These requirements include recognition of scale-type distinctions, unidimensionality, invariance, admissible arithmetic, dimensional homogeneity, and the principle that ordinal observations cannot be transformed into interval or ratio measures except under Rasch measurement rules.

The HTA knowledge space is defined pragmatically and operationally. Depending upon the institution or jurisdiction under examination, the corpus may include published HTA guidelines, agency decision frameworks, cost-effectiveness reference cases, academic journals and textbooks, modelling templates, technical reports, task-force recommendations, teaching materials, methodological articles, institutional white papers, and associated research publications. Collectively, these sources form the epistemic environment within which HTA practitioners develop their analytical frameworks, interpret evidence, and justify evaluative practices. The boundary of interrogation is therefore not the whole of medicine, economics, or public policy, but the specific textual ecosystem that sustains HTA reasoning.

The interrogation process relies on categorical endorsement probabilities. Endorsement does not imply that a statement appears verbatim within the corpus. Rather, it reflects the extent to which the concepts embodied in a statement are supported, reinforced, or implied by the knowledge base as a whole. A high endorsement probability indicates that a proposition is broadly consistent with the dominant conceptual structure of the corpus. A low endorsement probability indicates that the proposition is weakly represented, rarely articulated, absent, or contradicted by prevailing assumptions.

This distinction is crucial. The purpose is not to identify isolated references to representational measurement, ratio measurement, unidimensionality, Rasch measurement, or falsification. The purpose is to determine whether these concepts occupy a meaningful place within the intellectual environment. A single article mentioning unidimensionality does not imply that unidimensionality is an organizing principle of a discipline. Conversely, repeated reliance on utilities, QALYs, and simulation models may indicate strong endorsement of assumptions even when those assumptions are not explicitly stated.

The resulting endorsement profile therefore provides a measure of the conceptual structure of the knowledge base. Statements receiving consistently high endorsement identify concepts that are reinforced within the HTA environment. Statements receiving consistently low endorsement identify concepts that are neglected, marginalized, or absent. When propositions grounded in measurement science receive low endorsement while propositions embedded in the reference-case paradigm receive high endorsement, the resulting pattern is interpreted as evidence of measurement inversion. The interrogation therefore evaluates not individual opinions but the intellectual architecture of the HTA knowledge base itself.

CATEGORICAL ENDORSEMENT PROBABILITIES

In the present application, the interrogation is tightly bounded. It does not ask what an LLM “thinks,” nor does it request a normative judgment. Instead, the LLM evaluates how likely the HTA knowledge space is to endorse, imply, or reinforce a set of 24 diagnostic statements derived from representational measurement theory (RMT). Each statement is objectively TRUE or FALSE under RMT. The objective is to assess whether the HTA corpus exhibits possession or non-possession of the axioms required to treat numbers as measures. The interrogation creates a categorical endorsement probability: the estimated likelihood that the HTA knowledge base endorses the statement whether it is true or false; *explicitly or implicitly*.

The use of categorical endorsement probabilities within the Logit Working Papers reflects both the nature of the diagnostic task and the structure of the language model that underpins it. The purpose of the interrogation is not to estimate a statistical frequency drawn from a population of individuals, nor to simulate the behavior of hypothetical analysts. Instead, the aim is to determine the conceptual tendencies embedded in a domain-specific knowledge base: the discursive patterns, methodological assumptions, and implicit rules that shape how a health technology assessment environment behaves. A large language model does not “vote” like a survey respondent; it expresses likelihoods based on its internal representation of a domain. In this context, endorsement probabilities capture the strength with which the knowledge base, as represented within the model, supports a particular proposition. Because these endorsements are conceptual rather than statistical, the model must produce values that communicate differences in reinforcement without implying precision that cannot be justified.

This is why categorical probabilities are essential. Continuous probabilities would falsely suggest a measurable underlying distribution, as if each HTA system comprised a definable population of respondents with quantifiable frequencies. But large language models do not operate on that level. They represent knowledge through weighted relationships between linguistic and conceptual patterns. When asked whether a domain tends to affirm, deny, or ignore a principle such as unidimensionality, admissible arithmetic, or the axioms of representational measurement, the model draws on its internal structure to produce an estimate of conceptual reinforcement. The precision of that estimate must match the nature of the task. Categorical probabilities therefore provide a disciplined and interpretable way of capturing reinforcement strength while avoiding the illusion of statistical granularity.

The categories used, values such as 0.05, 0.10, 0.20, 0.50, 0.75, 0.80, and 0.85, are not arbitrary. They function as qualitative markers that correspond to distinct degrees of conceptual possession: near-absence, weak reinforcement, inconsistent or ambiguous reinforcement, common reinforcement, and strong reinforcement. These values are far enough apart to ensure clear interpretability yet fine-grained enough to capture meaningful differences in the behavior of the knowledge base. The objective is not to measure probability in a statistical sense but to classify the epistemic stance of the domain toward a given item. A probability of 0.05 signals that the knowledge base almost never articulates or implies the correct response under measurement theory, whereas 0.85 indicates that the domain routinely reinforces it. Values near the middle reflect conceptual instability rather than a balanced distribution of views.

Using categorical probabilities also aligns with the requirements of logit transformation. Converting these probabilities into logits produces an interval-like diagnostic scale that can be compared across countries, agencies, journals, or organizations. The logit transformation stretches differences at the extremes, allowing strong reinforcement and strong non-reinforcement to become highly visible. Normalizing logits to the fixed ± 2.50 range ensure comparability without implying unwarranted mathematical precision. Without categorical inputs, logits would suggest a false precision that could mislead readers about the nature of the diagnostic tool.

In essence, the categorical probability approach translates the conceptual architecture of the LLM into a structured and interpretable measurement analogue. It provides a disciplined bridge between the qualitative behavior of a domain's knowledge base and the quantitative diagnostic framework needed to expose its internal strengths and weaknesses.

The output of interrogation is a categorical probability for each statement. Probabilities are then transformed into logits $[\ln(p/(1-p))]$, capped to ± 4.0 logits to avoid extreme distortions, and normalized to ± 2.50 logits for comparability across countries. A positive normalized logit indicates reinforcement in the knowledge base. A negative logit indicates weak reinforcement or conceptual absence. Values near zero logits reflect epistemic noise.

Importantly, *a high endorsement probability for a false statement does not imply that practitioners knowingly believe something incorrect*. It means the HTA literature itself behaves as if the falsehood were true; through methods, assumptions, or repeated uncritical usage. Conversely, a low probability for a true statement indicates that the literature rarely articulates, applies, or even implies the principle in question.

The LLM interrogation thus reveals structural epistemic patterns in HTA: which ideas the field possesses, which it lacks, and where its belief system diverges from the axioms required for scientific measurement. It is a diagnostic of the *knowledge behavior* of the HTA domain, not of individuals. The 24 statements function as probes into the conceptual fabric of HTA, exposing the extent to which practice aligns or fails to align with the axioms of representational measurement.

INTERROGATION STATEMENTS

INTERROGATING THE CURRICULUM

The identification of measurement inversion across HTA research centers, reimbursement agencies and academic programs raises an obvious question: where does this inversion originate? If the same conceptual failures are observed repeatedly across institutions and jurisdictions, then the explanation cannot rest solely with individual researchers, policy analysts or decision makers. A more plausible explanation is that these failures reflect deficiencies in the educational environment that prepares future HTA practitioners. If concepts central to measurement science are absent from curriculum content, then their absence from research practice should not be surprising. This realization provides the rationale for interrogating the curriculum.

The objective of curriculum interrogation differs from that of previous HTA knowledge-based assessments. Earlier interrogations focused on whether institutions recognized the requirements of

representational measurement and the standards necessary for quantitative claims. Curriculum interrogation asks a different question. Are faculty, students and researchers exposed to the concepts necessary to understand and apply those standards? The focus shifts from methodological outputs to educational inputs. Rather than examining what faculty, students and researchers do, attention is directed to what they are taught and what they know.

The importance of this distinction should not be underestimated. Educational programs do not merely transmit technical skills. They define the conceptual framework through which future practitioners understand evidence, measurement and scientific inquiry. Concepts that are absent from the curriculum are unlikely to emerge spontaneously in research practice. Equally, concepts that are emphasized repeatedly become part of the intellectual assumptions that shape subsequent analysis. If measurement inversion is widespread, then one possibility is that the educational foundations required to recognize and avoid measurement inversion have never been systematically incorporated into HTA teaching and research training.

For this reason, the curriculum interrogation was designed around a series of canonical statements intended to identify the presence or absence of foundational measurement concepts. These statements were deliberately elementary. The purpose was not to assess advanced methodological knowledge but to determine whether faculty, students and researchers are likely to encounter the principles that underpin lawful quantitative claims. The resulting framework begins with the concept of an attribute as the object of measurement and proceeds through target attribute specification, scales of measurement, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement and falsifiable claims. Together, these statements define the minimum intellectual foundations required for a measurement-based approach to therapy assessment in education.

These statements are:

- **An attribute is the specific outcome of interest in a therapy assessment.**
- **Every therapy assessment begins with specification of the target attribute.**
- **The principal scales of measurement (nominal, ordinal, interval and ratio) have different properties and support different forms of analysis.**
- **The measurement status of a target attribute must be established before quantitative claims can be advanced.**
- **The axioms of representational measurement underpin quantitative claims.**
- **Attributes must be demonstrated to be unidimensional before measurement is possible.**
- **A manifest attribute is directly observable and capable of supporting empirical observation.**
- **A latent attribute is not directly observable and requires a measurement model to estimate possession of the attribute.**
- **Manifest and latent attributes require different forms of ratio measurement.**
- **Therapy impact claims must be falsifiable.**

These ten statements form a logical sequence:

Attribute → Target Attribute → Scales of Measurement → Measurement Status → Representational Measurement → Unidimensionality → Manifest Attribute → Latent Attribute → Ratio Measurement → Falsifiable Claims

Together they define the minimum curriculum content required for a measurement-based approach to HTA and provide the framework for evaluating curriculum coverage.

The categorical probabilities reported in this assessment are intended as indicators of the extent to which a concept is represented within the curriculum knowledge base. They should not be interpreted as precise statistical estimates but as measures of the likelihood that a student, researcher or professional exposed to that knowledge base would encounter, recognize and subsequently endorse the canonical statement. In practical terms, the probability reflects the visibility and prominence of a concept within the educational environment associated with a research center or policy agency.

A high probability indicates that the concept is well represented within curriculum materials, research outputs and educational activities and is therefore likely to be familiar to students and researchers. Conversely, a low probability suggests that the concept is absent, only weakly represented, or occupies a peripheral position within the curriculum knowledge base. Students exposed to such an environment would therefore be unlikely to recognize the concept as an important component of HTA education and practice.

The probabilities should be viewed comparatively rather than in isolation. Their principal value lies in identifying patterns of curriculum coverage across institutions and concepts. In particular, low probabilities associated with scales of measurement, representational measurement, unidimensionality and ratio measurement indicate that these topics are unlikely to form a substantial part of the educational experience of the average student. The resulting profile provides an indication of curriculum strengths, deficiencies and potential areas for reconstruction.

RESULTS: A CONSENSUS PATTERN OF CURRICULUM INVERSION

The results of the seven United States curriculum interrogations have profound implications for the future of health technology assessment (Table 1) The evidence demonstrates a remarkably consistent pattern of curriculum inversion. Foundational concepts required for quantitative science receive weak endorsement, while the educational environment continues to reinforce the analytical framework associated with utilities, QALYs, economic evaluation, and reference-case simulation modelling. The findings are strikingly uniform across institutions. The issue is therefore not one of local educational priorities or individual program design. Rather, it reflects the existence of a shared educational framework that has become detached from the principles governing scientific measurement.

The significance of these findings extends well beyond curriculum design. Curriculum inversion provides the missing explanation for the pattern of measurement inversion documented in the companion interrogation of the same research centers. If students, researchers, analysts, and policy makers are not exposed to representational measurement, scale theory, admissible arithmetic, unidimensionality, manifest and latent attributes, dimensional homogeneity, ratio

measurement, and Rasch measurement, then it becomes unsurprising that these concepts are largely absent from subsequent research practice. A discipline cannot be expected to recognize deficiencies in measurement if the concepts required to identify those deficiencies are absent from its educational foundations.

The relationship between curriculum inversion and measurement inversion is therefore direct. The former reproduces the latter. Students are taught how to construct utilities, calculate QALYs, undertake cost-effectiveness analyses, and build simulation models before they are taught the conditions under which quantitative claims are scientifically legitimate. Arithmetic precedes measurement in education just as it precedes measurement in research practice. The educational environment reproduces the analytical assumptions of the reference-case paradigm while neglecting the principles necessary to evaluate those assumptions.

The implications are serious. A curriculum characterized by inversion is not merely incomplete; it is fundamentally misleading. Students may graduate with considerable technical expertise in modelling and economic evaluation while lacking the conceptual framework necessary to determine whether the quantities entering those analyses satisfy the requirements of measurement. The result is a self-replicating intellectual system. Each generation inherits the assumptions of the previous one because the curriculum does not provide the conceptual tools necessary to challenge them.

Once curriculum inversion and measurement inversion are considered together, a more profound conclusion emerges. The problem is no longer confined to particular methods, models, or analytical techniques. The problem lies within the paradigm itself. The reference-case framework depends upon utilities, QALYs, cost-effectiveness ratios, and simulation models. Yet the educational environment that supports these constructs provides little recognition of the measurement principles required to justify them. The research environment exhibits the same pattern. What appears to be methodological sophistication therefore rests upon a foundation that lacks the conditions necessary for quantitative science.

This leads directly to the question of paradigm failure. Paradigms survive because they provide a coherent intellectual framework through which evidence is generated, interpreted, and defended. The findings reported here indicate that the contemporary HTA paradigm no longer satisfies that requirement. Its educational foundations exhibit curriculum inversion. Its research foundations exhibit measurement inversion. Together they reveal a framework in which arithmetic has displaced measurement, numerical construction has displaced measurement theory, and methodological convention has displaced scientific validation.

The consequence is unavoidable. The contemporary HTA paradigm cannot be sustained through incremental methodological refinement because the deficiencies lie in its educational and measurement foundations. Curriculum inversion generates measurement inversion; measurement inversion exposes paradigm failure. The challenge facing HTA is therefore not reform but reconstruction. The future of the discipline depends upon replacing a framework built upon arithmetic before measurement with one grounded in representational measurement, lawful ratio measurement, and evaluable claims regarding therapy impact. Only then can HTA claim to operate within the standards of normal science.

TABLE 1: CURRICULUM CONTENT ENDORSEMENT - US HTA RESEARCH CENTERS

CANONICAL STATEMENT	CATEGORICAL PROBABILITY ENDORSEMENT						
	SCHAEFFER	HARVARD	CHOICE	DUKE	LEONARD	BLOOMBERG	STANFORD
An attribute is the specific outcome of interest in a therapy assessment	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Every therapy assessment begins with specification of the target attribute	0.80	0.85	0.85	0.85	0.85	0.85	0.85
The principal scales of measurement (nominal, ordinal, interval and ratio) have different properties and support different forms of analysis	0.20	0.20	0.20	0.20	0.20	0.20	0.20
The measurement status of a target attribute must be established before quantitative claims can be advanced	0.15	0.15	0.15	0.15	0.15	0.15	0.15
The axioms of representational measurement underpin quantitative claims	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Attributes must be demonstrated to be unidimensional before measurement is possible	0.10	0.10	0.10	0.10	0.10	0.10	0.10
A manifest attribute is directly observable and capable of supporting empirical observation	0.50	0.55	0.55	0.55	0.55	0.55	0.55
A latent attribute is not directly observable and requires a measurement model to estimate possession of the attribute	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Manifest and latent attributes require different forms of ratio measurement	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Therapy impact claims must be falsifiable	0.80	0.80	0.80	0.85	0.85	0.85	0.85

SCHAEFFER: Schaffer Center for Health Policy and Economics, USC; HARVARD: Harvard Center for Health Decision Science; CHOICE: University of Washington CHOICE Institute; DUKE: Duke-Margolis Institute for Health Policy, Duke University; LEONARD: Leonard Davis Institute of Health Economics, University of Pennsylvania; BLOOMBERG: Bloomberg School of Public Health, Johns Hopkins University; STANFORD: Stanford University Health Policy Center

FROM CURRICULUM INVERSION TO MEASUREMENT INVERSION TO PARADIGM FAILURE

The results of the seven United States curriculum interrogations have profound implications for the future of health technology assessment. The evidence demonstrates a remarkably consistent pattern of curriculum inversion. Foundational concepts required for quantitative science receive weak endorsement, while the educational environment continues to reinforce the analytical framework associated with utilities, QALYs, economic evaluation, and reference-case simulation modelling. The findings are strikingly uniform across institutions. The issue is therefore not one of local educational priorities or individual program design. Rather, it reflects the existence of a shared educational framework that has become detached from the principles governing scientific measurement.

The significance of these findings extends well beyond curriculum design. Curriculum inversion provides the missing explanation for the pattern of measurement inversion documented in the companion interrogation of the same research centres. If students, researchers, analysts, and policy makers are not exposed to representational measurement, scale theory, admissible arithmetic, unidimensionality, manifest and latent attributes, dimensional homogeneity, ratio measurement, and Rasch measurement, then it becomes unsurprising that these concepts are largely absent from subsequent research practice. A discipline cannot be expected to recognize deficiencies in measurement if the concepts required to identify those deficiencies are absent from its educational foundations.

The relationship between curriculum inversion and measurement inversion is therefore direct. The former reproduces the latter. Students are taught how to construct utilities, calculate QALYs, undertake cost-effectiveness analyses, and build simulation models before they are taught the conditions under which quantitative claims are scientifically legitimate. Arithmetic precedes measurement in education just as it precedes measurement in research practice. The educational environment reproduces the analytical assumptions of the reference-case paradigm while neglecting the principles necessary to evaluate those assumptions.

The implications are serious. A curriculum characterized by inversion is not merely incomplete; it is fundamentally misleading. Students may graduate with considerable technical expertise in modelling and economic evaluation while lacking the conceptual framework necessary to determine whether the quantities entering those analyses satisfy the requirements of measurement. The result is a self-replicating intellectual system. Each generation inherits the assumptions of the previous one because the curriculum does not provide the conceptual tools necessary to challenge them.

Once curriculum inversion and measurement inversion are considered together, a more profound conclusion emerges. The problem is no longer confined to particular methods, models, or analytical techniques. The problem lies within the paradigm itself. The reference-case framework depends upon utilities, QALYs, cost-effectiveness ratios, and simulation models. Yet the educational environment that supports these constructs provides little recognition of the measurement principles required to justify them. The research environment exhibits the same

pattern. What appears to be methodological sophistication therefore rests upon a foundation that lacks the conditions necessary for quantitative science.

This leads directly to the question of paradigm failure. Paradigms survive because they provide a coherent intellectual framework through which evidence is generated, interpreted, and defended. The findings reported here indicate that the contemporary HTA paradigm no longer satisfies that requirement. Its educational foundations exhibit curriculum inversion. Its research foundations exhibit measurement inversion. Together they reveal a framework in which arithmetic has displaced measurement, numerical construction has displaced measurement theory, and methodological convention has displaced scientific validation.

The consequence is unavoidable. The contemporary HTA paradigm cannot be sustained through incremental methodological refinement because the deficiencies lie in its educational and measurement foundations. Curriculum inversion generates measurement inversion; measurement inversion exposes paradigm failure. The challenge facing HTA is therefore not reform but reconstruction. The future of the discipline depends upon replacing a framework built upon arithmetic before measurement with one grounded in representational measurement, lawful ratio measurement, and evaluable claims regarding therapy impact. Only then can HTA claim to operate within the standards of normal science.

DOES A NEW PARADIGM HAVE A FUTURE

Judged by the combined findings for curriculum inversion and measurement inversion, the conclusion appears unavoidable. The contemporary HTA paradigm has reached the point where its scientific legitimacy can no longer be taken for granted. The educational foundations of the discipline fail to provide students and researchers with the concepts necessary to evaluate quantitative claims, while the analytical framework itself relies upon utilities, QALYs, cost-effectiveness ratios, and reference-case simulation models that fail to satisfy the standards of representational measurement. Taken together, these findings point not merely to methodological weakness but to paradigm failure.

The implications are profound. If the critique presented here is accepted, then the future of HTA cannot lie within the existing framework. Utilities, QALYs, and reference-case simulation models are not peripheral components of contemporary HTA; they are its defining features. Once their measurement foundations are rejected, the framework loses the analytical instruments upon which it depends. The logical consequence is that HTA must transition to a new paradigm grounded in lawful measurement, with linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes providing the basis for evaluable and falsifiable claims regarding therapy impact.

Whether such a transition will occur is a different question. Scientific arguments do not operate in a vacuum. The reference-case paradigm has dominated HTA for more than forty years. Entire careers have been built around utility measurement, QALY construction, economic modelling, and cost-effectiveness analysis. Academic departments, research centers, journals, professional societies, consulting organizations, and government agencies have invested heavily in the existing framework. Thousands of publications, methodological guidelines, doctoral theses, and

reimbursement decisions are linked to its assumptions. For many individuals and institutions, acceptance of paradigm failure would require acknowledging that a substantial part of their professional activity was built upon concepts whose measurement status was never adequately established.

This creates a powerful force for resistance. History suggests that established paradigms rarely disappear because their supporters voluntarily abandon them. More commonly, they persist through institutional inertia, methodological refinement, and attempts to defend the status quo. It would therefore be unsurprising if many HTA organizations respond to the present critique by proposing revised utility instruments, improved modelling techniques, enhanced methodological guidance, or more sophisticated simulation methods. Yet none of these responses address the central issue. The problem is not insufficient methodological complexity. The problem is the absence of lawful measurement.

At the same time, circumstances have changed. The findings for measurement inversion and curriculum inversion are now public and reproducible. Students, faculty members, manufacturers, policy makers, and research sponsors can no longer assume that the measurement foundations of the reference-case paradigm are secure. Questions that were once rarely asked can now be posed directly. Where is representational measurement taught? Why are ratio measurement and Rasch measurement absent from the curriculum? How can QALYs be defended if their measurement properties remain unproven? The intellectual environment has changed because awareness has changed.

For this reason, the future of HTA may ultimately be determined less by the willingness of established institutions to change than by the willingness of new entrants to accept the old framework. Once curriculum inversion and measurement inversion have been identified, the burden of proof shifts to those who wish to preserve the status quo. The challenge is no longer to demonstrate that the reference-case paradigm has failed. The challenge is to explain why it should continue. In the long run, that may prove a far more difficult task.

To support the transition to a new paradigm Maimon Research as developed a nine-unit HTA Reconstruction Program that provides a comprehensive educational framework for replacing the traditional reference-case paradigm with one grounded in representational measurement and the standards of normal science. Beginning with the principle that measurement must precede arithmetic, the program introduces attributes, scale theory, ratio measurement, manifest and latent attributes, and the role of Rasch logit ratio measurement for latent outcomes. It demonstrates why utilities, QALYs, and reference-case simulation models fail the requirements for lawful quantitative claims and presents an alternative framework based on evaluable and falsifiable assessments of therapeutic impact. Designed for students, faculty, researchers, manufacturers, and HTA agencies, the program serves as a practical resource for curriculum reconstruction and the transition to measurement-based health technology assessment. Details are given on the Maimon Research website ¹¹ .

CONCLUSIONS

The findings reported in this assessment leave little room for ambiguity. Across seven leading United States HTA research centers, the interrogation results reveal a stable and remarkably consistent pattern of measurement inversion. The foundational principles required for quantitative science receive weak endorsement, while the assumptions that sustain utilities, QALYs, cost-effectiveness ratios, and reference-case simulation models receive strong reinforcement. This pattern is not confined to a single institution, methodological tradition, or research group. It is a consensus profile that characterizes the contemporary US HTA knowledge base.

The significance of these findings extends far beyond methodological criticism. The issue is not whether existing HTA methods require refinement or whether additional statistical sophistication might improve current practice. The issue is that the measurement foundations necessary to support quantitative claims have never been secured. Concepts such as representational measurement, ratio measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, manifest and latent attributes, and Rasch measurement remain largely absent from the intellectual framework within which HTA operates. The consequence is that arithmetic has been allowed to assume the role that properly belongs to measurement. Numerical constructions are treated as measures, while the requirements that distinguish measures from numbers remain largely unexamined.

The evidence also points to a deeper problem. Measurement inversion does not emerge spontaneously. It is reproduced through curriculum inversion. Students, researchers, analysts, and policy makers are trained within educational environments that emphasize utility assessment, QALY construction, cost-effectiveness analysis, and simulation modelling while giving little attention to the principles required to determine whether these activities are scientifically legitimate. The result is a self-sustaining system in which the assumptions of the reference-case paradigm are transmitted from one generation to the next without exposure to the conceptual tools needed to challenge them. The persistence of the paradigm therefore reflects institutional reproduction rather than scientific validation.

The implications for HTA are profound. Once measurement inversion and curriculum inversion have been identified, there can be no return to the status quo. The contemporary framework cannot be rescued through revised utility instruments, updated methodological guidance, more sophisticated simulation models, or increasingly elaborate statistical techniques. None of these responses address the underlying problem. Arithmetic cannot create measurement where measurement does not exist. The challenge is therefore not one of reform but of reconstruction.

Fortunately, a path forward exists. A scientifically credible HTA framework begins with attributes, distinguishes manifest from latent outcomes, recognizes that measurement must precede arithmetic, and restricts quantitative claims to lawful ratio measures. This leaves only two admissible measurement frameworks: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. Together they provide the foundation for evaluable, replicable, and falsifiable claims regarding therapy impact. The nine-unit Maimon Research LLC reconstruction program offers a practical transformation resource through which these principles can be incorporated into research, education, and professional practice.

The choice facing HTA is therefore stark. One option is to continue defending a paradigm whose central constructs fail the standards required for quantitative science. The other is to acknowledge the implications of measurement inversion and curriculum inversion and undertake the transition to a framework grounded in lawful measurement. The findings reported here suggest that the future credibility of HTA will depend upon that choice. If reconstruction is embraced, HTA may yet emerge as a discipline consistent with the standards of normal science. If it is resisted, the relevance and legitimacy of the reference-case paradigm will continue to erode as awareness of its measurement deficiencies becomes increasingly widespread.

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 United States: The absence of representational measurement and the Schaeffer Center for Health Policy and Economics, University of Southern California. Logit Working Paper 750. <https://maimonresearch.com/logit-working-paper-no-750-april-2026/>
- ² United States: The absence of representational measurement and the Center for Health Decision Science, Harvard University. Logit Working Paper No 751. <https://maimonresearch.com/logit-working-paper-no-751-april-2026/>
- ³ United States: The absence of representational measurement and the University of Washington CHOICE Institute. Logit Working Paper No 752. <https://maimonresearch.com/logit-working-paper-no-752-april-2026/>
- ⁴ United States: The absence of representational measurement and the Duke-Margolis Institute for Health Policy, Duke University. Logit Working Paper No 753. <https://maimonresearch.com/logit-working-paper-no-753-april-2026/>
- ⁵ United States: The absence of representational measurement and the Leonard Davis Institute of Health Economics, University of Pennsylvania. Logit Working Paper No 754. <https://maimonresearch.com/logit-working-paper-no-754-april-2026/>
- ⁶ United States: The absence of representational measurement and the Health Economics Group, Bloomberg School of Public Health. Johns Hopkins University. Logit Working Paper No 756. <https://maimonresearch.com/logit-working-paper-no-756-april-2026/>
- ⁷ United States: The absence of representational measurement and the Stanford University Health Policy Center. Logit working Paper No 755. <https://maimonresearch.com/logit-working-paper-no-755-april-2026/>

⁸ Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

⁹ Krantz D, Luce R, Suppes P, Tversky A. Foundations of Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

¹⁰ Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed). New York Routledge, 2021

¹¹ Maimon Research Transition Program <https://maimonresearch.com/hta-reconstruction-program-and-fees/>