

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



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**CANADA: PARADIGM FAILURE IN HEALTH
TECHNOLOGY ASSESSMENT**

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

LOGIT WORKING PAPER No 545 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper presents a comprehensive assessment of paradigm failure in Canadian health technology assessment (HTA). The argument advanced is that the contemporary reference-case framework fails not because of methodological shortcomings or the need for further refinement, but because it was established without satisfying the standards required for quantitative science. Drawing upon a series of large language model (LLM) interrogations of Canadian HTA knowledge bases, the paper demonstrates the presence of both measurement inversion and curriculum inversion across agencies, research centers, professional organizations, journals, and educational environments. The institutions examined include Canada's Drug Agency (CDA), the Canadian Journal of Health Technologies (CJHT), Health Technology Assessment International (HTAi), the Health Utilities Index (HUI), and five leading Canadian HTA research centers: THETA, HTPU, IHE, ICES, and C2E2.

The findings reveal a remarkably consistent pattern. Concepts fundamental to representational measurement—including scales of measurement theory, ratio measurement, dimensional homogeneity, unidimensionality, manifest and latent attributes, admissible arithmetic, and Rasch measurement—receive little recognition. At the same time, utilities, QALYs, cost-effectiveness ratios, and simulation models are accepted as legitimate quantitative constructs despite lacking demonstrated measurement properties. The result is a self-reinforcing intellectual framework in which arithmetic precedes measurement and numerical constructions are routinely treated as evidence. The Canadian curriculum interrogations further demonstrate that students and researchers are trained to apply reference-case methods while receiving little exposure to the principles required to evaluate their scientific legitimacy. Curriculum inversion therefore emerges as the educational mechanism through which measurement inversion is reproduced across successive generations of HTA practitioners.

Particular attention is given to the Health Utilities Index, which is presented as an example of a broader cascade of measurement failure extending from health-state valuations to utility weights, QALYs, and simulation-based value claims. The paper also highlights the near-complete absence of Rasch measurement and the failure to distinguish manifest from latent attributes, despite the central role of latent outcomes in contemporary HTA. The conclusion is that the reference-case paradigm represents a forty-year design failure. Reconstruction is therefore required. A future HTA framework must be grounded in representational measurement, distinguish manifest from latent attributes, employ linear ratio measures for manifest outcomes and Rasch logit ratio measures for latent outcomes, and require all claims regarding therapy impact to be evaluable, replicable, and capable of falsification. Only through such a transition can HTA establish a scientifically credible foundation for quantitative claims regarding therapeutic value.

INTRODUCTION

Paradigm failure has not suddenly emerged within HTA. The conditions for failure were present from the beginning. What is new is not the failure itself but the evidence. The recent knowledge-base interrogations, including those undertaken in Canada, provide systematic evidence that the assumptions underpinning the reference-case paradigm never satisfied the standards of representational measurement required for quantitative claims.

The Canadian results demonstrate a consistent failure to recognize the requirements of representational measurement, admissible arithmetic and scientific measurement theory. Across agencies, research centers and curriculum environments there is little appreciation that measurement must precede arithmetic, that scale properties constrain mathematical operations, and that quantitative claims require demonstrated measurement rather than numerical construction. The significance of these findings extends beyond individual methods or institutions. They expose a discipline that has largely failed to recognize the standards necessary to support quantitative claims regarding therapy impact.

The issue is therefore not whether HTA has recently lost its scientific footing. The issue is that it never secured that footing in the first place. The reference-case paradigm was built upon assumptions regarding utilities, QALYs and simulation modelling that were accepted without demonstrating the measurement properties necessary to support them. What the interrogations reveal is not the collapse of a previously valid framework but the exposure of a framework that never satisfied the requirements of quantitative science.

Canada has long regarded itself as occupying a leading position in international health technology assessment (HTA). Through influential research centers, national organizations, methodological guidance and university training programs, it has played a prominent role in promoting the reference-case framework as the preferred approach to therapy assessment. Utility scores, quality-adjusted life years (QALYs), cost-effectiveness ratios and simulation models have become established not merely as analytical tools but as defining features of modern HTA. The widespread belief has been that the reference case represents the culmination of methodological progress in healthcare decision making.

Nothing could be further from the truth.

Judged against the standards of representational measurement, the reference case is an analytically impossible construct. Its central components depend upon arithmetic operations that require measurement properties which have never been demonstrated. Utility scores are treated as though they possess ratio properties. QALYs are created through multiplication without establishing that multiplication is admissible. Reference-case simulation models extend these assumptions across hypothetical populations and lifetime horizons. The framework is therefore grounded not in measurement but in assumptions regarding measurement.

The significance of this conclusion is often misunderstood. The problem is not that the reference case was once scientifically defensible and has subsequently deteriorated. Nor is it a matter of methodological refinement failing to keep pace with advances in science. The defect was present from the beginning. The moment health-state valuations were transformed into utility scores and treated as measurable quantities, the framework departed from the requirements of measurement. QALYs, cost-effectiveness ratios and simulation models merely extended and institutionalized that original error. The reference case did not begin with measurement and later lose its way. It began without measurement. The train never left the station.

This raises an obvious question. If the defect was present from the outset, why did it remain largely invisible for more than four decades? Why was it accepted by researchers, policy makers,

educators and professional organizations? Why was a framework founded on inadmissible arithmetic able to establish itself as the dominant paradigm in HTA?

The answer emerges from the Canadian LLM interrogations and the broader international interrogation program. Their significance lies not simply in documenting measurement failure but in explaining why it persisted. The concepts required to identify the defect are themselves largely absent from the HTA knowledge base. Across agencies, research centers and curriculum environments there is little recognition of representational measurement, admissible arithmetic, ratio measurement, unidimensionality, the distinction between manifest and latent attributes, or the role of Rasch measurement in constructing measures of latent attribute possession. The result is a professional culture in which arithmetic is accepted while measurement is assumed. The analytical impossibility at the heart of the reference case remained invisible because the intellectual tools required to identify it were largely absent.

The Canadian findings are therefore important for reasons that extend well beyond Canada itself. For the first time, the interrogations provide systematic and reproducible evidence that the persistence of the reference case is not evidence of scientific validity. Rather, it reflects a curriculum and research environment that has institutionalized measurement inversion and reproduced it through teaching, research and policy for more than forty years. What has been exposed is not a temporary methodological weakness but a paradigm founded on assumptions that never satisfied the standards required for quantitative science.

The implications are unavoidable. Reconstruction is no longer optional. HTA must be rebuilt upon a foundation that recognizes measurement as the prerequisite for quantitative claims, distinguishes manifest from latent attributes, employs valid ratio measures where required and utilizes Rasch measurement for latent attributes. Claims regarding therapy impact must be evaluable, replicable and capable of falsification. Evidence must be grounded in measurement rather than numerical construction. The conclusion is stark but unavoidable: the reference case represents a forty-year design failure. Its future lies not in reform but in replacement.

CANADA: DEFINING PARADIGM FAILURE IN HTA

It may seem odd to describe HTA as experiencing paradigm failure when the evidence presented in this report indicates that the field never satisfied the measurement standards required to support quantitative claims in the first place. In one sense, the failure was present from the outset. The reference-case paradigm was built upon assumptions regarding utilities, QALYs and simulation modelling that were accepted without demonstrating the measurement properties necessary to support them.

Yet from the perspective of the many thousands of researchers, policy makers, educators and analysts who accepted and applied the reference case over the past four decades, a paradigm clearly existed. It shaped research agendas, guided reimbursement decisions, informed university curricula and defined professional standards. Its assumptions were rarely questioned because they were treated as established knowledge. The reference case acquired the status of orthodoxy.

What is new is not the failure itself but the evidence. The recent knowledge-base, measurement inversion and curriculum inversion interrogations provide systematic and reproducible evidence that the assumptions underpinning the reference case never possessed the scientific foundations that were taken for granted. The paradigm did not suddenly collapse; rather, its absence of foundations has now been exposed. In this sense, paradigm failure has become visible. What has failed is not a once-valid framework but a belief system whose analytical weaknesses remained hidden because the concepts required to identify them were largely absent from the discipline itself. The interrogations therefore mark the point at which the intellectual legitimacy of the reference-case paradigm can no longer be sustained.

One of the most curious features of HTA is that the reference-case paradigm was never designed to generate evaluable claims. From its inception, the framework emphasized the construction of models, projections and hypothetical estimates rather than claims capable of empirical assessment, replication and potential falsification. In most scientific disciplines, the absence of evaluable claims would have raised immediate concerns because the growth of objective knowledge depends upon the ability to confront claims with evidence. Yet within HTA this absence passed largely unnoticed.

Instead of emphasizing evaluation and replication, the field increasingly focused on validation. Models were validated against assumptions, expert opinion, internal consistency checks and comparisons with other models. The question of whether the claims generated by those models could be empirically evaluated in real-world populations received far less attention. Validation became a substitute for falsification. As a consequence, HTA evolved into a discipline where the appearance of analytical rigor displaced the scientific requirement that claims be capable of being proven wrong. A closed system to create threshold driven cost-effectiveness claims.

This distinction is fundamental. Science advances through evaluable and falsifiable claims, not through the construction of increasingly sophisticated frameworks whose outputs remain insulated from empirical challenge. The reference case therefore departed not only from the standards of measurement science but also from the principles governing the evolution of objective knowledge. Its central products were designed to support decisions, not to generate claims that could be tested, replicated and potentially refuted. The result was a methodological culture in which numerical credibility was often mistaken for scientific credibility. What the recent interrogations reveal is that this departure from evaluability was not a peripheral feature of the reference case. It was built into its design from the beginning.

Paradigm failure occurs when the conceptual foundations of a framework are shown to be defective. The framework may continue to function administratively, but it can no longer claim scientific authority. Applied to HTA, paradigm failure rests upon two related failures. The first is internal. The central constructs of the reference-case paradigm violate the requirements of measurement scales and the axioms of representational measurement^{1 2 3}. Utilities are treated as though they possess ratio properties without demonstration. QALYs are created through multiplication without establishing that multiplication is admissible. Reference-case simulation models extend these assumptions across hypothetical populations and lifetime horizons. The result is a framework that depends upon arithmetic operations which the underlying quantities cannot

support. The failure is therefore not one of calculation but of measurement. The quantities entering the analysis do not possess the properties required for the arithmetic imposed upon them.

The second failure is external. The axioms of representational measurement are not merely technical requirements governing arithmetic. Their purpose is to support empirical inquiry by creating measures that allow quantitative claims to be evaluated against observation. Measurement exists to support the growth of objective knowledge through evaluable, replicable and potentially falsifiable claims. Yet the reference-case paradigm was never designed to generate claims of this type. Its outputs are modelled projections, hypothetical lifetime estimates and simulated outcomes that cannot be subjected to direct empirical assessment. Rather than emphasizing evaluation and falsification, the paradigm evolved around validation, where assumptions are compared with other assumptions and models are assessed against other models.

Taken together, these failures are decisive. Internally, the reference case lacks the measurement foundations required to support its arithmetic. Externally, it lacks a framework for generating evaluable and falsifiable claims. The paradigm therefore fails both tests that define scientific inquiry. It cannot support lawful quantitative claims and it cannot support the systematic growth of objective knowledge. What has been exposed through the recent interrogation studies is not simply a methodological weakness but a framework whose foundations are incompatible with the requirements of both measurement science and empirical science.

CANADA HTA PARADIGM FAILURE: THE EVIDENCE FOR MEASUREMENT INVERSION

Measurement inversion was assessed through a structured large language model (LLM) interrogation framework designed to determine whether the foundational requirements of representational measurement were recognized within a given HTA knowledge base. The interrogation employed a fixed set of 24 canonical statements derived from the axioms of measurement theory. These statements addressed concepts such as unidimensionality, admissible scale types, dimensional homogeneity, ratio measurement, latent and manifest attributes, the precedence of measurement over arithmetic, and the requirements for evaluable and falsifiable claims. Of the 24 statements, 14 were designated as scientifically correct (TRUE) and 10 as scientifically incorrect (FALSE). The objective was not to assess opinion, but to determine the extent to which the knowledge base endorsed propositions that were either consistent or inconsistent with the principles of representational measurement.

For each statement, the LLM was asked to assess whether the proposition was supported by the knowledge base under examination. Responses were recorded as categorical probabilities ranging from very low endorsement to very high endorsement. These probabilities reflected the degree to which the proposition appeared to be accepted, embedded, or reinforced within the published materials, methodological guidance, educational content, and associated documentation of the institution or organization. A probability close to 1.0 indicated strong endorsement, while a probability close to 0.0 indicated strong rejection. The resulting probability profile provided a structured representation of the knowledge base and its alignment with measurement principles.

Measurement inversion was identified when scientifically correct statements received consistently low probabilities of endorsement while scientifically incorrect statements received consistently high probabilities. In practical terms, this meant that propositions such as “measurement must precede arithmetic,” “multiplication requires ratio scale properties,” and “latent attributes require explicit measurement models” were rejected or weakly supported, while propositions such as “utilities are ratio measures,” “QALYs represent valid quantitative outcomes,” or “simulation models generate evaluable evidence” were strongly endorsed. The resulting pattern provided evidence of a systematic inversion of measurement principles in which arithmetic operations and numerical constructions were accepted without first establishing the measurement properties required to support them. The analysis therefore focused not on isolated methodological errors but on the overall structure of the knowledge base and the extent to which it institutionalized a framework of measurement inversion.

The evidence for measurement inversion and paradigm failure in Canada is considered from a number of perspectives: first, from a national level; second, the knowledge bases of leading HTA research centers; third, the knowledge base of the Health Utility Instruments (HUI); fourth, the knowledge base of Health Technology Assessment International (HTAi) and, finally, the knowledge base of the Canadian Journal of Health Technologies (CJHT).

The results of the national assessment demonstrated that measurement inversion is not an isolated methodological weakness within Canadian health technology assessment (HTA), but a pervasive and institutionalized characteristic ⁴. The assessment found little evidence that measurement functions as a governing constraint on analytical practice. Fundamental concepts such as unidimensionality, admissible scale types, dimensional homogeneity, invariance, and the requirement that measurement must precede arithmetic were either absent or only weakly represented. At the same time, utilities derived from preference instruments were routinely treated as though they possessed interval or ratio properties, negative utility values were accepted within purported quantitative frameworks, and QALYs were assumed to support arithmetic operations involving aggregation, multiplication and cost-effectiveness comparisons. The distinction between manifest and latent attributes was rarely recognized, while Rasch measurement principles the only framework capable of constructing quantitative measures for latent attributes from ordinal observations were largely absent from the Canadian corpus.

The resulting endorsement profile was neither mixed nor transitional. Rather, it revealed a stable and internally coherent epistemic structure in which numerical constructions are accepted as evidence without first satisfying the requirements for measurement. The Canadian HTA community has therefore not merely overlooked representational measurement; it has developed an analytical framework that operates independently of it. The consequence is that many of the quantitative claims generated by contemporary Canadian HTA lack the measurement foundations necessary to support evaluable, replicable and scientifically credible assessments of therapeutic value.

The assessment of Canada's leading health technology assessment (HTA) research centers produced results that closely mirrored those observed for Canada's Drug Agency (CDA), the Canadian Journal of Health Technologies (CJHT) and the broader national HTA knowledge base ^{5 6 7}. The centers examined included the Toronto Health Economics and Technology Assessment

Collaborative (THETA) at the University Health Network and University of Toronto ⁸, the Health Technology and Policy Unit (HTPU) at the University of Alberta ⁹, the Institute of Health Economics (IHE) in Edmonton ¹⁰, the Community Health and Epidemiology research group at Dalhousie University ¹¹, the Centre for Clinical Epidemiology and Evaluation (C2E2) at University of British Columbia ¹², and health economics and outcomes research programs associated with ICES ¹³. Despite differences in institutional history, research focus, and organizational structure, the interrogation results were remarkably consistent ¹⁴.

Across all centers there was no evidence that the principles of representational measurement functioned as foundational constraints on research design, methodological development, or the interpretation of value claims. Statements affirming the core requirements of measurement theory received uniformly low levels of endorsement. These included propositions that measurement must precede arithmetic, that multiplication requires ratio scale properties, that latent and manifest attributes require different measurement approaches, that dimensional homogeneity is a prerequisite for quantitative claims, and that unidimensionality must be demonstrated before numerical operations can be justified. Across centers, these propositions consistently occupied the lower end of the endorsement spectrum.

Conversely, propositions inconsistent with representational measurement received strong support. Utility scores derived from multiattribute preference instruments were treated as though they possessed interval or ratio scale properties. QALYs were accepted as valid quantitative representations of therapeutic value. Simulation models were assumed capable of generating meaningful long-term estimates of treatment impact despite relying upon utility-based inputs whose measurement status was not examined. The arithmetic manipulation of utility scores, the aggregation of heterogeneous dimensions into composite indices, and the multiplication of utility values by time were routinely accepted as methodological norms. These practices were not presented as assumptions requiring validation but as established elements of HTA methodology.

Particularly notable was the near-complete absence of Rasch measurement theory and the broader literature on latent variable measurement. The interrogations found little recognition that patient-reported outcomes, quality-of-life constructs, symptom burden, functional status, and similar attributes are latent phenomena requiring explicit measurement models if quantitative claims are to be supported. Instead, ordinal responses were commonly treated as though they could be transformed into quantitative evidence through scoring algorithms, index construction, or preference weighting. The distinction between ordinal ranking and quantitative measurement was largely absent from the knowledge base.

The consistency of these findings is important. The results do not suggest methodological diversity, active debate, or competing measurement paradigms within Canadian HTA research. Rather, they indicate the presence of a stable and self-reinforcing intellectual framework in which the axioms of representational measurement are effectively excluded from methodological practice. The probability–logit profiles generated from the interrogations were strikingly similar demonstrating that the observed pattern was not institution-specific but characteristic of the Canadian HTA research environment as a whole.

The overall conclusion is that Canada's leading HTA research centers have institutionalized a framework of measurement inversion. Numerical outputs are routinely generated, modelled, aggregated, and interpreted without first establishing that the underlying attributes possess the measurement properties required to support those operations. Rather than providing an independent scientific challenge to prevailing HTA assumptions, these centers largely function as mechanisms through which those assumptions are reproduced, legitimized, and transmitted throughout Canadian HTA research, teaching, and policy development.

CANADA: MEASUREMENT INVERSION AND THE HEALTH UTILITIES INDEX (HUI)

The absence of awareness identified in the curriculum interrogations extends beyond individual research centers. This is evident in the continuing prominence of the Health Utilities Index (HUI), one of Canada's most influential contributions to utility measurement and health outcomes assessment¹⁵. For decades, the HUI has been presented as a legitimate source of utility values for economic evaluation, reimbursement decisions and QALY construction. Yet the development and application of the HUI proceeded without addressing the fundamental requirements of representational measurement. The assumption that health-state valuations could be transformed into utility scores possessing the properties required for arithmetic operations was largely accepted rather than demonstrated. As a consequence, one of the foundational instruments of Canadian HTA became embedded within the reference-case framework despite lacking the measurement foundations necessary to support the quantitative claims subsequently derived from it.

The significance of the HUI extends far beyond the instrument itself. Its acceptance illustrates how deeply measurement inversion became institutionalized within Canadian HTA. The issue is not simply that the HUI is employed in economic evaluations. The issue is that its widespread acceptance reflects a broader failure to recognize the standards that govern measurement. Questions regarding scale properties, ratio measurement, dimensional homogeneity, admissible arithmetic and the distinction between measurement and numerical assignment were largely absent from the surrounding knowledge base. The result is that the same failure identified in Canadian research centers and curriculum environments is reproduced at the national level. The HUI stands not merely as a utility instrument but as an example of how an entire HTA tradition evolved without confronting the measurement requirements upon which its quantitative claims ultimately depended.

The HUI also illustrates a broader cascade of measurement failure that characterizes contemporary HTA. The process begins with the valuation of health-state descriptions, where preferences are elicited from respondents and treated as though they represent measurable quantities. These valuations are then converted through scoring algorithms into utility weights, despite the absence of evidence that the resulting scores possess ratio properties. The utility weights are subsequently employed as inputs to the QALY, where they are multiplied by survival time under the assumption that multiplication is admissible. Finally, QALYs are incorporated into cost-effectiveness models and reference-case simulations to generate claims regarding value, efficiency, and resource allocation. At each stage, arithmetic proceeds as though the required measurement properties have already been established. Yet the central question is never resolved: do the quantities entering the analysis satisfy the requirements of ratio measurement? The failure to address this question means

that the defect is cumulative rather than isolated. The HUI is therefore not simply a multiattribute utility instrument. It is one component in a chain of assumptions that extends from health-state valuation to utility construction, QALY generation, and simulation-based value claims. The entire chain rests upon unproven assumptions regarding measurement.

CANADA: MEASUREMENT INVERSION AND HEALTH TECHNOLOGY ASSESSMENT INTERNATIONAL (HTAi)

Health Technology Assessment International (HTAi) is the principal global professional organization for health technology assessment. Founded in 2003, it brings together researchers, policy-makers, clinicians, industry representatives, patient groups, and HTA agencies from around the world. Its purpose is to promote the development and application of HTA in healthcare decision-making through conferences, policy forums, special interest groups, educational activities, and professional networking. HTAi does not itself undertake reimbursement decisions or formal technology assessments. Rather, it serves as an international forum for the exchange of ideas, methods, and best practices in HTA. Through its annual meetings, working groups, publications, and collaborations with national HTA agencies, HTAi has played a major role in disseminating and legitimizing the reference-case framework, including the use of utilities, QALYs, cost-effectiveness analysis, and simulation modelling as standard components of contemporary HTA practice.

HTAi helps define what constitutes acceptable HTA practice across the world. It serves not merely as a professional society but as an important source of epistemic authority. The methods, assumptions and analytical frameworks endorsed within HTAi are disseminated internationally through research centers, government agencies, academic programs and policy organizations. If the standards of representational measurement were embedded within the intellectual foundations of HTA, one would expect them to be visible within the knowledge environment created and sustained by HTAi.

The interrogation evidence suggests otherwise¹⁶. The HTAi knowledge base exhibits the same pattern observed in national agencies, academic research centers and university curricula. The principles that define quantitative science receive little recognition, while propositions that conflict with those principles receive strong endorsement. Measurement does not occupy its required position as the foundation of quantitative claims. Instead, arithmetic is treated as self-evidently legitimate and the measurement properties necessary to support that arithmetic are assumed rather than demonstrated. Concepts such as representational measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, ratio measurement and Rasch measurement are largely absent from the methodological framework promoted by the organization.

This finding is particularly significant because HTAi operates at the level where international standards are discussed, disseminated and normalized. The issue is therefore not whether particular agencies or research centers have misunderstood measurement theory. The issue is that the international body responsible for promoting methodological leadership in HTA appears to share the same assumptions. Preference-based utilities are treated as though they possess interval or ratio properties. QALYs are accepted as legitimate products of multiplication. Negative values are accommodated within purported ratio scales. Simulation models are assumed to generate

meaningful quantitative evidence. Yet the measurement foundations required to support these practices remain largely unexamined.

The absence of Rasch measurement is especially revealing. Across many scientific disciplines, Rasch measurement has become the accepted framework for constructing measures of latent attribute possession from ordinal observations. Given the central importance of patient-reported outcomes, quality of life, functioning, symptoms and related latent attributes within HTA, one might expect Rasch principles to occupy a prominent place within the international methodological discourse. Their near absence indicates that the problem extends beyond individual instruments or analytical techniques. It reflects a broader failure to recognize the distinction between numerical scoring and measurement.

The significance of the HTAi interrogation therefore lies not merely in the endorsement profile itself but in what it reveals about the international HTA community. The findings suggest that measurement inversion has become institutionalized at the highest levels of methodological authority. The same assumptions identified in Canadian agencies, research centers and curriculum environments are reproduced internationally through the very organization charged with advancing HTA science. This is powerful evidence that the problem is not local, national or organizational. It is paradigmatic.

From this perspective, HTAi provides perhaps the clearest demonstration of why measurement inversion has proven so durable. The concepts required to challenge the reference-case paradigm are largely absent from the institutions responsible for defining methodological standards. As a result, the reference case continues to be promoted as best practice even though its foundational assumptions fail the requirements of representational measurement. The consequence is an international HTA culture that has normalized non-measurement while presenting itself as a quantitative science. What the recent interrogations reveal is that this position can no longer be sustained. The issue is not methodological disagreement but paradigm failure at the level where HTA standards themselves are defined.

MEASUREMENT INVERSION: *CANADIAN JOURNAL OF HEALTH TECHNOLOGIES (CJHT)*

It is also important to consider the knowledge base represented by the *Canadian Journal of Health Technologies (CJHT)* ¹⁷. This is significant because CJHT serves as the official publication platform of Canada's Drug Agency (CDA), formerly CADTH, and functions as the principal vehicle through which the Agency disseminates the evidence and analyses that support health technology assessment and reimbursement decision-making in Canada.

The interrogation concluded that the CJHT operates as a centralized repository for Canadian HTA practice. The journal publishes a wide range of assessments, including full health technology assessments, rapid reviews, reimbursement evaluations, horizon scans, patient and stakeholder reports, and methodological discussions. These publications are intended to inform decisions regarding the adoption, funding and appropriate use of pharmaceuticals, medical devices and clinical interventions within publicly funded health systems.

The journal's orientation is policy-facing rather than purely academic. Its content is designed to provide decision-makers with synthesized evidence on clinical effectiveness, safety and economic impact, frequently incorporating cost-effectiveness analyses, utility-based outcome measures and simulation modelling frameworks. In doing so, it reflects and reinforces the methodological assumptions that underpin contemporary Canadian HTA practice, including the use of multiattribute utility instruments, quality-adjusted life years, and long-term modeled projections of therapy value.

As a consequence, CJHT functions as more than a publication outlet. It acts as a codified expression of the conceptual and methodological foundations of Canadian HTA. The journal provides an institutional record of the assumptions that are accepted as evidence, the forms of analysis regarded as legitimate, and the standards that guide reimbursement recommendations and policy decisions.

From the perspective of representational measurement, however, the interrogation identified the same limitations that characterize the wider Canadian HTA environment. The journal's knowledge base assumes the legitimacy of utility scores, QALYs and simulation-based economic models without recognizing the prior requirement that claims involving arithmetic operations must be supported by admissible measurement scales. There is no recognition that measurement must precede arithmetic, that multiplication requires ratio scale properties, that latent attributes require explicit measurement models, or that dimensional homogeneity is a prerequisite for quantitative claims.

The significance of this finding extends beyond the journal itself. Because CJHT is the formal publication platform of CDA, it serves as a mechanism through which methodological assumptions are reproduced, legitimized and transmitted throughout the Canadian HTA community. The absence of representational measurement standards within this knowledge base therefore contributes to the continuing institutionalization of measurement inversion. Rather than functioning as a forum in which the foundations of quantitative claims are critically examined, the journal largely reflects and reinforces an existing framework in which numerical constructions are accepted as evidence without first satisfying the requirements for measurement.

Consequently, any transition of Canadian HTA to a scientifically defensible measurement framework must extend beyond reimbursement agencies and academic centers to include the publication standards embodied in the Canadian Journal of Health Technologies. Unless the journal requires explicit demonstration of the measurement properties that support quantitative claims, it will continue to serve as a vehicle through which the assumptions of the current paradigm are preserved and reproduced. The result is the perpetuation of a policy framework in which arithmetic is routinely applied in the absence of measurement, and where the appearance of quantitative rigor substitutes for the standards required by normal science.

CANADA: THE EXTENT OF MEASUREMENT INVERSION

For Canada, the overall result is a self-reinforcing system in which the reference case is both taught and applied while the measurement principles necessary to evaluate its scientific legitimacy remain largely absent. The issue is therefore not confined to a particular method or institution. It extends across policy, research and professional education.

The critical point is that paradigm failure does not imply that the framework ceases to exist. Paradigms often survive long after their foundations have been undermined. Educational programs continue to teach them. Research centers continue to apply them. Agencies continue to use them in decision making. What changes is their intellectual status. They become institutional conventions rather than scientific frameworks.

Eventually, however, a choice must be made. Either the standards of measurement science are abandoned in order to preserve the paradigm, or the paradigm is abandoned in order to preserve the standards of measurement science. In the case of HTA, the relevant standards are those of representational measurement, dimensional homogeneity, unidimensionality, ratio measurement and falsifiable claims. If these standards are accepted, then utilities, QALYs and reference-case simulation models cannot be defended in their present form.

Paradigm failure therefore marks the point at which reform becomes impossible. The deficiencies are no longer matters of refinement or methodological adjustment. The foundations themselves have failed. Reconstruction becomes necessary because the existing structure cannot be repaired without abandoning the assumptions upon which it was built.

For Canada, this conclusion applies equally to the reference-case framework employed by the CDA and to the educational and research structures that continue to reproduce it. The issue is not whether the current paradigm can be improved; it cannot. The issue is whether a framework constructed on arithmetic before measurement can survive once measurement is restored to its central place in scientific inquiry; it cannot. What is being observed is not methodological weakness but paradigm failure.

The implications are unavoidable. The future of HTA in Canada cannot be secured through incremental reform, revised guidelines or increasingly sophisticated simulation techniques. The problem lies deeper, in the assumptions upon which the reference-case framework itself depends. Once those assumptions are shown to be incompatible with the requirements of measurement, the only remaining option is reconstruction. The task is not to repair the paradigm but to replace it with a framework in which measurement precedes arithmetic, quantitative claims are evaluable and falsifiable, and evidence is grounded in measures rather than numerical constructions.

This finding is important because measurement inversion lies at the heart of the contemporary failed HTA paradigm. Measurement inversion occurs when arithmetic precedes measurement. Instead of establishing the measurement properties of a quantity and then determining which mathematical operations are admissible, HTA begins with arithmetic and assumes that measurement has already been achieved. Numbers are treated as measures because they are numerical. The scientific burden of demonstrating measurement is quietly abandoned.

The national interrogation of HTA in Canada demonstrated this pattern clearly. The knowledge base gave strong endorsement to propositions that support utility-based assessment, QALY construction and reference-case simulation modelling, while showing little recognition of the principles of representational measurement. The result was a profile consistent with those observed in Australia, the United States, Canada and Europe. Numerical constructions were accepted as though they represented valid measures despite the absence of evidence that the underlying quantities possessed the properties required for arithmetic operations.

CANADA: PARADIGM FAILURE - THE EVIDENCE FOR CURRICULUM INVERSION

Curriculum inversion occurs when a curriculum teaches the application of quantitative methods while failing to teach the measurement principles that determine whether those methods are scientifically legitimate. In a scientifically coherent curriculum, measurement precedes arithmetic. Students first learn the nature of attributes, the requirements of representational measurement, the distinctions among nominal, ordinal, interval and ratio scales, and the conditions necessary for valid quantitative claims. Only then are they introduced to the arithmetic, statistical and modelling procedures that depend upon those measurement properties. Curriculum inversion reverses this sequence. Students learn how to calculate, model and analyze before they learn how to determine whether the quantities entering those analyses are measures. Arithmetic becomes detached from measurement and numerical manipulation is treated as though it were equivalent to quantitative science.

The consequences are profound. A curriculum affected by inversion reproduces a professional culture in which measurement is assumed rather than demonstrated. Concepts such as unidimensionality, dimensional homogeneity, admissible arithmetic, manifest and latent attributes, ratio measurement and Rasch measurement either disappear entirely or are treated as peripheral concerns. Students become proficient in the techniques of economic evaluation, utility assessment, QALY construction and simulation modelling without acquiring the conceptual tools necessary to evaluate the legitimacy of those methods. The result is that the curriculum not only fails to identify measurement errors but actively reproduces them across successive generations of researchers, analysts and decision makers. Curriculum inversion therefore serves as the educational mechanism through which measurement inversion becomes institutionalized within a discipline.

For this reason, curriculum assessment emerges as a critical component of HTA reconstruction. The objective is not simply to determine whether students are exposed to contemporary HTA methods. Rather, it is to determine whether they are exposed to the foundational concepts that make the evaluation of those methods possible. A curriculum that emphasizes modelling, economic evaluation and decision analysis while neglecting measurement theory will inevitably reproduce the same conceptual limitations observed in current HTA practice.

The Canadian curriculum interrogation occupies a particularly important position within the broader assessment of health technology assessment (HTA) knowledge bases¹⁸. Australia, New Zealand, the United Kingdom, and other jurisdictions all demonstrate evidence of measurement inversion. Canada, however, provides an opportunity to examine the educational mechanisms through which measurement inversion is reproduced and sustained. The interrogation of five

leading Canadian HTA research centers, THETA, HTPU, IHE, ICES, and C2E2, offers a unique perspective on the relationship between education, research, policy formation, and the continued acceptance of the reference-case paradigm.

The results are striking not because they reveal large differences between institutions, but because they reveal almost none. Despite differences in institutional history, geographic location, research focus, and organizational structure, the endorsement profiles are remarkably uniform. Across all five centers the same concepts receive strong endorsement, the same concepts receive weak endorsement, and the same omissions recur. This consistency is arguably the most important finding of the entire assessment because it points to the existence of a shared intellectual framework rather than isolated curriculum deficiencies.

The strongest endorsements are associated with outcome identification and target specification. The proposition that an attribute is the specific outcome of interest in a therapy assessment receives probabilities ranging from 0.75 to 0.80, while the proposition that every therapy assessment begins with specification of the target attribute receives probabilities ranging from 0.65 to 0.75. These findings indicate that Canadian HTA training places considerable emphasis on identifying outcomes and defining research questions. Students and researchers are taught to determine what is to be evaluated and how interventions are to be compared. In this respect, the curriculum performs exactly as one would expect within a mature research environment.

The difficulty is that the curriculum largely stops there. Once outcomes have been identified, the interrogation reveals an abrupt collapse in endorsement for the concepts that transform outcomes into measures. Across all five centers, the proposition that the principal scales of measurement possess different properties and support different forms of analysis receives a probability of only 0.20. The proposition that the measurement status of a target attribute must be established before quantitative claims can be advanced receives a probability of only 0.15. The proposition that representational measurement underpins quantitative claims receives a probability of only 0.05 across every center. These are among the lowest endorsement profiles observed in any curriculum interrogation.

The implications are profound. Outcome specification is not measurement. Identifying an attribute merely establishes what is to be examined. It does not establish whether the attribute possesses the properties necessary to support arithmetic operations or quantitative claims. Yet the Canadian curriculum appears to move directly from outcome identification to analysis while largely bypassing the question of measurement. Students are taught what to evaluate but not how to determine whether the quantities entering the evaluation qualify as measures. Arithmetic is therefore detached from measurement at the educational level long before it appears in research practice.

The comparison with Australia and New Zealand is instructive. Australian interrogations of seven HTA research centers revealed a similar pattern. Strong recognition was given to outcomes and evaluation, while representational measurement, ratio measurement, latent attributes, and Rasch measurement remained largely absent. New Zealand interrogations of PHARMAC, the University of Auckland, and the University of Otago produced almost identical findings. In all three countries the curriculum environment displays a common structure: students are introduced to utilities,

QALYs, cost-effectiveness analysis, and simulation modelling without first being equipped to evaluate the measurement foundations upon which those methods depend. Canada therefore does not stand alone. Rather, it confirms the existence of a broader international pattern of curriculum inversion. What distinguishes Canada is the degree of consistency with which the pattern appears across multiple institutions and multiple knowledge bases.

Particularly revealing is the treatment of latent attributes. Across all five Canadian centers, the proposition that latent attributes require a measurement model receives a probability of only 0.10. Equally low is the proposition that manifest and latent attributes require different forms of ratio measurement. This latter proposition receives a probability of only 0.05 across every institution. These findings suggest that one of the most important distinctions in measurement science is effectively absent from Canadian HTA education. Students may learn to evaluate survival, resource utilization, and adverse events alongside pain, functioning, quality of life, and patient satisfaction without recognizing that these represent fundamentally different classes of attributes requiring fundamentally different measurement frameworks.

This omission has major implications for the place of Rasch measurement within Canadian HTA. If latent attributes are not explicitly recognized, there is little reason to introduce the measurement framework required for their assessment. The near absence of latent attribute concepts therefore helps explain the near absence of Rasch measurement. The result is that students and researchers are exposed to utility instruments and patient-reported outcome measures without being introduced to the only established framework capable of constructing invariant measures of latent attribute possession. The educational environment therefore reproduces the same conceptual omissions identified previously in Canadian HTA practice.

This relationship becomes even more significant when viewed alongside the broader Canadian HTA landscape. The curriculum findings mirror those observed in interrogations of Canada's Drug Agency, the Canadian Journal of Health Technologies, and the Health Utilities Index. The concepts absent from educational environments are the same concepts absent from policy guidance, research publications, and assessment frameworks. The HUI provides a particularly striking example. For decades it has occupied a central place within Canadian outcomes research and economic evaluation. Yet the curriculum interrogation helps explain why its measurement foundations have rarely been subjected to sustained scrutiny. Students are not systematically exposed to representational measurement, ratio measurement, dimensional homogeneity, or latent variable measurement. Consequently, the transformation of health-state valuations into utility scores, the construction of QALYs, and the use of these quantities in simulation models appear unproblematic. The educational framework reproduces the assumptions required to sustain the broader HTA paradigm.

The significance of these findings extends beyond curriculum design. They provide a plausible explanation for the persistence of measurement inversion itself. If students are not taught representational measurement, they cannot be expected to apply it. If they are not taught the distinction between manifest and latent attributes, they cannot be expected to recognize the different measurement requirements associated with each. If they are not taught Rasch measurement, they cannot be expected to challenge utility-based approaches to latent outcomes.

Curriculum inversion therefore becomes the mechanism through which measurement inversion is transmitted from one generation of HTA practitioners to the next.

Viewed in this context, the Canadian curriculum interrogation provides more than evidence of educational deficiencies. It demonstrates how an intellectual framework reproduces itself. The same omissions appear in teaching, research, policy analysis, journal publication, and technology assessment. The concepts required to challenge the reference-case paradigm are largely absent from the environments responsible for training those who operate within it. Canada therefore provides one of the clearest demonstrations yet available that curriculum inversion and measurement inversion are not separate phenomena. They are complementary features of the same paradigm. Together they explain how a framework built upon utilities, QALYs, and simulation models has persisted despite its failure to satisfy the standards of measurement science.

CANADA: UNDERSTANDING RATIO MEASUREMENT

The fatal weakness of contemporary health technology assessment (HTA) in Canada is not a failure of modelling, statistical analysis or computational sophistication. It is a failure to understand ratio measurement¹⁹. This may appear an unlikely criticism given the technical complexity of modern HTA, yet it lies at the heart of the paradigm failure documented throughout this report. Once the requirements imposed by ratio measurement are understood, the intellectual foundations of utilities, QALYs and reference-case simulation models become impossible to defend.

Ratio measurement occupies a unique position within science. A ratio scale possesses a true zero and supports all arithmetic operations, including multiplication and division. The importance of this property cannot be overstated. Multiplication is not a universal mathematical operation that can be applied to any numbers that happen to be available. Multiplication is admissible only when the quantities involved possess the measurement properties required to support proportional interpretation. If these properties are absent, the arithmetic result may be a number, but it is not a measure.

This principle is recognized throughout the physical sciences. No scientist would multiply quantities drawn from arbitrary scales and claim that the result represented a meaningful measure. The scale properties of the quantities must first be established. Measurement precedes arithmetic.

The tragedy of the reference-case paradigm is that this principle was never applied to the quantities upon which the framework depends. The architects of utility-based assessment assumed that preference scores could function as ratio measures. Health-state valuations obtained from time trade-off, standard gamble and related exercises were treated as though they represented proportions of health. A utility score of 0.8 came to be interpreted as 80 percent of full health, while a score of 0.5 was interpreted as 50 percent of full health. Yet no demonstration was provided that these values possessed the properties required for proportional interpretation.

This failure proved decisive because the entire QALY framework depends upon multiplication. Survival time is unquestionably a ratio measure. To multiply survival time by a utility score, the utility score must itself function as a ratio measure. If it does not, the operation is inadmissible.

The resulting QALY is not a measure of quality-adjusted survival but a numerical construction generated through unlawful arithmetic.

The consequences extend even further. Reference-case simulation models depend upon QALYs as their central outcome measure. If the QALY lacks measurement status, every estimate generated by the model inherits the same defect. The complexity of the simulation is irrelevant. Sophisticated mathematics cannot rescue quantities that fail the requirements of measurement. Arithmetic cannot create measurement where measurement does not exist.

What makes this failure particularly significant in Canada is that the curriculum interrogations indicate little recognition of ratio measurement or the scale theory upon which it depends. Students are exposed to utilities, QALYs and simulation models but are not systematically introduced to the distinctions between nominal, ordinal, interval and ratio scales. They are not taught that different scales support different forms of arithmetic. Most importantly, they are not apparently taught that multiplication requires ratio measurement.

The consequence is predictable. Graduates learn how to calculate QALYs without learning whether QALYs are scientifically possible. They learn how to populate simulation models without learning whether the quantities entering those models qualify as measures. The scientific challenge is never encountered because the curriculum does not provide the concepts necessary to formulate it.

This omission helps explain why the reference-case paradigm remains largely unchallenged. If students are never introduced to the requirements of ratio measurement, they have no basis for questioning utility scores, QALYs or simulation outputs. The arithmetic appears legitimate simply because it is numerical. Measurement inversion becomes normalized because the measurement foundations are absent from professional education.

Yet once ratio measurement is restored to its central place, the situation changes dramatically. The entire structure of utility-based assessment depends upon a proposition that has never been demonstrated: that utility scores possess ratio properties. If this proposition fails, the QALY fails. If the QALY fails, reference-case simulation models fail. The collapse is not partial. It is complete because each component depends upon the one that precedes it.

This is why ratio measurement occupies such a critical place in the argument for paradigm failure. The issue is not whether utilities can be refined, whether valuation methods can be improved or whether simulation models can be made more sophisticated. The issue is whether the quantities entering those analyses satisfy the requirements of ratio measurement. If they do not, the resulting arithmetic is inadmissible regardless of the complexity of the analytical framework.

Understanding ratio measurement therefore changes everything. It reveals that the central weakness of the HTA paradigm is not methodological but conceptual. The failure to recognize the constraints imposed by ratio measurement undermines utilities, destroys the QALY and removes the foundation upon which reference-case simulation modelling depends. Once this is understood, the conclusion is unavoidable. The issue confronting HTA in Australia is not reform but

reconstruction. Ratio measurement does not merely challenge the reference case; it exposes why the reference case cannot survive. A design fault that guaranteed rejection.

CANADA: MANIFEST AND LATENT ATTRIBUTES

The importance of ratio measurement has been further obscured by the failure of contemporary HTA to distinguish between manifest and latent attributes. Although these represent fundamentally different classes of outcomes, both require ratio measurement if quantitative claims regarding therapy impact are to be scientifically defensible. The crucial point is that the form of ratio measurement differs according to the nature of the attribute being assessed.

For manifest attributes, the measurement problem is relatively straightforward. Manifest attributes are directly observable. Survival time, hospital admissions, emergency department visits, medication possession, treatment discontinuation, adverse events, and resource utilization can all be observed, counted, or timed directly. When appropriately specified, these attributes support linear ratio measures characterized by a meaningful zero and admissible arithmetic operations. The measurement task is therefore one of observation and quantification.

Latent attributes present an entirely different challenge. Pain, fatigue, depression, anxiety, physical functioning, treatment satisfaction, need fulfilment, and quality of life are not directly observable. Their existence must be inferred from observable indicators such as questionnaire responses or behavioral observations. Consequently, latent attributes cannot be measured through direct observation alone. They require a measurement model capable of transforming ordinal observations into a quantitative estimate of attribute possession.

This is where Rasch assumes a unique importance. Rasch measurement provides the only established framework capable of constructing a lawful measure of latent attribute possession while simultaneously testing whether the requirements for measurement have been satisfied. Through the conjoint calibration of persons and items, Rasch establishes a latent continuum, evaluates unidimensionality, tests invariance, and determines whether observations support quantitative interpretation. The resulting logit scale is not merely another scoring system. It is a measurement framework specifically designed for latent attributes.

The significance of this distinction for HTA is profound. Had the discipline been forced from the outset to classify every target attribute as either manifest or latent, attention would inevitably have focused on the corresponding measurement requirements. Investigators would have been required to identify whether a proposed outcome demanded direct ratio measurement or a latent measurement model. The measurement problem would have become unavoidable. The present reliance on utilities, preference scores, and composite indices would have been difficult to sustain because the question would immediately arise: what is the attribute, and how is it being measured?

Instead, contemporary HTA largely bypassed this distinction. Outcomes with fundamentally different measurement properties were incorporated into a common evaluative framework. Directly observable events, patient-reported outcomes, preference scores, and utility algorithms were treated as though they occupied the same measurement status. Once this occurred, the need

to distinguish between linear ratio measures and Rasch logit ratio measures disappeared from view. The measurement problem was effectively replaced by a scoring problem.

The consequence was that Rasch measurement never assumed the central role it should have occupied within HTA. If latent attributes are recognized as requiring their own form of ratio measurement, Rasch becomes indispensable. If the distinction between manifest and latent attributes is ignored, Rasch appears optional, peripheral, or unnecessary. The history of HTA reflects the latter path. The discipline embraced utilities, QALYs, and simulation models while largely overlooking the one framework capable of providing lawful measurement for latent outcomes.

Had HTA adopted the manifest-latent distinction at the outset, the history of the discipline would have been very different. Utilities would have been immediately challenged as measures of latent attributes. QALYs would have been challenged as inadmissible arithmetic. Rasch measurement would have occupied a central position in the assessment of patient-reported outcomes. The contemporary reference case might never have emerged in its present form.

Recognition of the manifest-latent distinction therefore changes the entire structure of HTA. It redirects attention from numerical constructions to measurement, from scores to attributes, and from preference weights to possession of measurable outcomes. Most importantly, it establishes that there are only two legitimate pathways to quantitative assessment: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. Once this distinction is accepted, the absence of Rasch from contemporary HTA is revealed not as a minor methodological omission but as one of the defining features of the paradigm's measurement failure.

CANADA: THE ABSENCE OF RASCH

One of the most striking findings from the Canadian measurement and curriculum interrogation is the near-complete absence of Rasch ratio measurement and its role in the assessment of latent attributes. This omission is critical in defining the paradigm failure of contemporary HTA. important because it reveals a fundamental weakness in the educational and methodological framework that underpins contemporary health technology assessment. The issue is not whether the term "Rasch" appears occasionally in conference abstracts, research presentations, or specialist publications. The issue is whether Rasch measurement is recognized as the essential framework for constructing quantitative measures of latent attributes. The interrogations say that it is not.

This omission is particularly significant because HTA places considerable emphasis on patient-centered outcomes, quality of life, symptom burden, functional status, treatment satisfaction, patient experience, and similar constructs. These are all latent attributes. They cannot be directly observed in the same way that hospital admissions, survival time, medication possession, or adverse events can be observed. Latent attributes exist, but they are not directly measurable through counting, timing, or simple observation. Their measurement requires a formal ratio measurement model.

This is where Rasch occupies a unique position. Rasch is not simply another psychometric technique competing with item response theory, PROMIS, utility instruments, or preference-based

scoring systems. Rasch addresses a fundamentally different question. It asks whether ordinal observations can be transformed into a quantitative measure of possession of a latent attribute. In doing so, it provides the only established framework capable of demonstrating whether the conditions required for measurement have been satisfied for latent attributes.

The distinction is critical. Patient-reported outcomes typically begin with ordinal responses to questionnaire items. Patients may indicate levels of pain, fatigue, anxiety, mobility limitations, or functional difficulties. These responses are rankings. They provide information about order but not quantity. Arithmetic performed directly on ordinal observations cannot create measurement. Summing scores, averaging responses, applying weights, or generating utility algorithms does not transform ordinal observations into quantitative measures. Numerical manipulation is not measurement.

The Rasch model was developed in the 1950s precisely to address this problem. Through the conjoint calibration of persons and items, Rasch analysis estimates the location of respondents on a latent continuum while simultaneously testing whether the data satisfy the requirements for ratio measurement. Unidimensionality, invariance, item fit, category functioning, local independence, and differential item functioning are not optional refinements. They are the conditions that must be satisfied before claims regarding possession of a latent attribute can be advanced. Rasch therefore provides both a measurement model and a set of empirical tests for determining whether measurement is possible.

The interrogations make clear that this perspective is absent from the Australian HTA educational framework. Students and practitioners are introduced to patient-reported outcomes, utility instruments, preference weights, quality-of-life measures, and value assessment methodologies without first confronting the measurement problem those constructs are intended to address. The curriculum appears to move directly from patient responses to scoring systems and economic evaluation. The intermediate step, demonstrating that a latent attribute has been measured, is effectively bypassed.

This omission has important consequences. Without Rasch ratio measurement, latent attributes remain latent. Utility scores, composite indices, and preference-weighted algorithms may generate numerical outputs, but they do not establish that the underlying construct has been measured. The existence of a number should not be confused with the existence of a measure. Yet much of contemporary HTA proceeds as though this distinction does not matter.

The result is that students are trained to accept numerical representations of quality of life, patient benefit, symptom burden, and treatment impact without being introduced to the Rasch framework required to determine whether those representations possess measurement properties. They learn how utilities are generated, how QALYs are constructed, and how economic models are populated, but they are not taught how latent attributes can be measured. The educational sequence is therefore inverted. Numerical outputs are presented before the conditions required to justify those outputs.

The absence of Rasch is consequently more than a methodological omission. It is a defining characteristic of curriculum inversion. The curriculum recognizes the importance of latent attributes but fails to recognize the only framework capable of transforming observations of those

attributes into quantitative ratio measures. This leaves faculty, students and practitioners with a vocabulary of scores, utilities, and indices but without an understanding of measurement itself. Until Rasch measurement assumes its proper place within HTA education, latent attributes will continue to be represented through numerical constructions rather than lawful measures, and the distinction between scoring and measurement will remain obscured.

CANADA: THE HISTORICAL SPECTACLE

One of the most remarkable aspects of the HTA paradigm is not the nature of its failure but its longevity. Errors occur in every scientific discipline. Hypotheses are proposed, tested, challenged and, when necessary, abandoned. This process is fundamental to scientific progress. What makes the history of HTA unusual, if not unique, is that a framework that failed the requirements of measurement became established as the dominant model for evaluating therapy impact and remained largely unchallenged for more than four decades at a global level.

The scale of this enterprise is difficult to overstate. Since the emergence of utility-based assessment in the late twentieth century, thousands, and perhaps tens of thousands, of papers have been published employing utility scores, QALYs, cost-effectiveness ratios and reference-case simulation models. Entire academic careers have been devoted to the development and refinement of these methods (e.g., the latest PubMed hit for QALYS yields 28,870 responses). Research centers have been established to advance them. Professional societies have promoted them. Governments have incorporated them into reimbursement decision making. Universities have embedded them within curricula and professional training programs. An extensive international infrastructure emerged to support and sustain the reference-case paradigm.

Viewed from within the discipline, this growth was often interpreted as evidence of success. The expanding literature, increasing methodological sophistication and widespread institutional adoption appeared to demonstrate intellectual maturity. Yet institutional acceptance and scientific validity are not the same thing. The crucial question was never how many studies employed the framework but whether the framework itself satisfied the standards required for quantitative claims.

The historical spectacle lies in the fact that this question remained largely unasked. Utility scores were accepted as though they possessed ratio properties. QALYs were accepted as though multiplication was admissible. Simulation models were accepted as though they generated quantitative evidence. At every stage, arithmetic was treated as self-justifying. The prior requirement that measurement be demonstrated before arithmetic could proceed was largely ignored.

This is not a criticism of individual investigators. Most researchers worked within the intellectual environment they inherited. The concepts, methods and assumptions of the reference case were presented as established knowledge. Young investigators learned them from textbooks, mentors and professional training programs. Manuscripts employing these methods passed through peer review. Research grants supported further methodological development. Professional advancement often depended upon participation in the framework rather than its criticism.

The consequence was the creation of a self-reinforcing intellectual system or, in the terminology of Richard Dawkins, a memplex. Authors wrote papers based upon accepted assumptions. Reviewers evaluated submissions using the same assumptions. Editors selected manuscripts from a pool of work operating within the same conceptual boundaries. Research centers trained students using the same methods. Policy agencies relied upon the resulting analyses. The framework reproduced itself because the standards used to evaluate it were derived from the framework itself.

The significance of the recent measurement and curriculum interrogations is that they provide, for the first time, a systematic explanation for how this occurred. The persistence of the reference case cannot be explained simply by professional inertia or institutional conservatism. The interrogations reveal something more fundamental. The concepts required to identify the problem are themselves largely absent from the HTA knowledge base. Representational measurement, admissible arithmetic, dimensional homogeneity, unidimensionality, ratio measurement, the distinction between manifest and latent attributes and the role of Rasch measurement occupy little place in the curriculum, research and policy structures of the discipline.

This finding is important because it shifts the discussion from methodological criticism to paradigm failure. The problem is not merely that HTA adopted a flawed framework. The problem is that the intellectual tools required to recognize the flaw were largely excluded from the discipline. Researchers could not easily challenge assumptions regarding measurement because the principles necessary to formulate that challenge were absent from their professional education and research environment.

The result is one of the most striking episodes in the history of applied research. An emerging discipline adopted an analytical framework whose central constructs failed the requirements of measurement. Rather than examining those foundations, successive generations of researchers devoted their efforts to elaborating the framework. More sophisticated utility instruments were developed. More complex simulation models were constructed. Larger datasets were assembled. Increasingly advanced analytical techniques were employed. Yet the fundamental question remained untouched: are the quantities entering these analyses measures?

This is not unprecedented. The history of science contains examples of theories and practices that persisted for long periods despite foundational weaknesses. What distinguishes HTA is that the persistence of the paradigm was reinforced by the absence of the very concepts required to challenge it. The framework became self-reinforcing because measurement itself disappeared from view. As a consequence, criticism focused on methodological refinement rather than on the more fundamental question of whether the framework could satisfy the requirements of quantitative science.

From the perspective of measurement science, the answer is clear. The issue is not that the reference case occasionally produced misleading results. The issue is that its central constructs never satisfied the conditions necessary to support the arithmetic upon which they depended. The historical spectacle is therefore not merely the persistence of error. It is the persistence of error on an extraordinary scale.

Future historians of HTA may regard this period as a cautionary example of how scientific disciplines can become detached from their own foundations. The spectacle is not that the methods lacked sophistication. The spectacle is that increasing sophistication concealed an elementary failure. For more than forty years, an immense intellectual effort was devoted to refining arithmetic while neglecting measurement. The result was not the accumulation of quantitative knowledge but the institutionalization of measurement inversion.

That is why the present moment represents more than a methodological dispute. It marks the point at which an entire analytical tradition must confront the standards it neglected. The historical spectacle is coming to an end. What remains is the task of reconstruction.

CANADA: MEETING PROFESSIONAL STANDARDS

The evidence presented in this paper points to a conclusion that can no longer be avoided. The professional standards required to evaluate health technologies scientifically are not present in the Canadian HTA curriculum knowledge base. This is not a matter of incomplete coverage or a need for modest curriculum adjustment. The omission concerns the foundations of quantitative science itself. Students and researchers are introduced to the reference case, utilities, QALYs, economic evaluation and simulation modelling, but they are not systematically introduced to the measurement principles necessary to determine whether these constructs can support valid quantitative claims.

This distinction is decisive. A professional curriculum should not simply train students to operate within an existing framework. It should equip them to evaluate whether that framework satisfies the standards of scientific inquiry. In HTA, that means understanding attributes, scales of measurement, representational measurement, unidimensionality, dimensional homogeneity, manifest and latent attributes, Rasch measurement, ratio measurement and falsifiable claims. These are not optional topics. They are the intellectual tools required to distinguish a measure from a number, lawful arithmetic from inadmissible arithmetic, and evidence from numerical construction. Without them the claim to provide a professional training falls short.

A further concern is the extent to which measurement inversion has become embedded within the Canadian HTA environment. Institutional traditions that have been accepted for decades are rarely abandoned quickly, particularly when they have shaped professional education, research priorities, reimbursement policies, and methodological standards. For more than thirty years, the Health Utilities Index (HUI) has been regarded as a legitimate foundation for utility assessment and QALY construction. Similarly, HTAi, CADTH, and now the CDA have played central roles in promoting and disseminating the reference-case framework throughout Canada and internationally. These organizations did not merely participate in the development of the paradigm; they helped establish the intellectual environment within which it became accepted as standard practice.

The consequences are evident in the research-centered interrogations. The same patterns of curriculum inversion and measurement inversion appear repeatedly across institutions, indicating that the underlying assumptions of the reference-case framework have been transmitted largely without challenge. Concepts such as representational measurement, ratio measurement,

unidimensionality, dimensional homogeneity, manifest and latent attributes, and Rasch measurement remain weakly represented or absent. The result is a self-reinforcing system in which successive generations of researchers, educators, and policy analysts inherit the same conceptual limitations as their predecessors.

If Canada is to move toward a new paradigm for HTA, the task will therefore be substantial. The challenge is not merely methodological but institutional. It requires replacing a framework that has become deeply embedded within teaching, research, and policy with one grounded in the standards of measurement science. Yet the existence of measurement inversion and curriculum inversion does not mean that transition is impossible. Rather, it demonstrates why transition is necessary. The longer these assumptions remain unchallenged, the more difficult reform becomes. The findings reported here suggest that the future credibility of Canadian HTA will depend upon its willingness to confront those assumptions directly and to replace them with a framework built upon lawful measurement, evaluable claims, and the standards of normal science.

The consequence is professionally serious. A lifetime commitment to the reference case is no longer defensible once the requirements of measurement are recognized. The reference case is not a neutral analytical tool awaiting refinement. It is a sequence of dependent errors. Time trade-off responses are treated as proportional values without demonstrating ratio properties. Health-state valuations are converted into utility weights without establishing measurement status. Utilities are multiplied by survival time without proving that multiplication is admissible. QALYs are then inserted into simulation models that project lifetime claims using quantities whose measurement foundations remain unproven. One error follows another.

These errors remain largely invisible so long as scales of measurement and representational measurement are absent from professional training. If students are never taught that different scales support different arithmetic operations, they cannot see why utility scores may not support multiplication. If they are never taught that ratio measurement requires a true zero, they cannot see why bounded utility values do not automatically function as proportions. If they are never taught dimensional homogeneity, they cannot see why multiplying ordinal or interval preference scores by ratio time is inadmissible. If they are never taught the distinction between manifest and latent attributes and the contribution of Rasch measurement, they cannot see why quality of life requires a different measurement framework from survival time or hospitalization.

This is why curriculum reconstruction is unavoidable. The status quo curriculum is unsustainable because it reproduces the very paradigm that has failed. It trains students to accept utilities, QALYs and reference-case simulation models while withholding the measurement framework that would expose their weaknesses. This is not professional preparation. It is professional enculturation into a defective paradigm.

The required alternative is already clear. A reconstructed curriculum must begin with measurement. Students must first learn that every therapy assessment concerns an attribute. They must then distinguish manifest from latent attributes. Manifest attributes, such as survival time, hospital admissions, emergency department visits and treatment persistence, may support direct linear ratio measurement. Latent attributes, such as pain, fatigue, functional status, need fulfilment and quality of life, require a measurement model capable of estimating possession of the attribute.

Rasch measurement provides the necessary framework for constructing measures of latent attribute possession. Only after measurement has been demonstrated can arithmetic and claims assessment proceed.

A professional HTA curriculum must therefore move from simulation method training to scientific training. It must teach students how to construct evaluable claims, specify measurement models, determine scale properties, assess admissible arithmetic and frame claims capable of empirical falsification. The objective should not be proficiency in the reference case. The objective should be competence in determining whether any proposed analytical framework satisfies the requirements of measurement and normal science.

A MANAGED TRANSITION TO MEASUREMENT

The results reported throughout this assessment cannot be dismissed as isolated findings or methodological disagreements. Interrogations of Canada's Drug Agency, the Canadian Journal of Health Technologies, leading HTA research centers, and the wider Canadian HTA literature all point to the same conclusion: the foundational requirements of representational measurement are absent from the Canadian HTA knowledge base. The evidence is now sufficiently extensive that the issue can no longer be regarded as a matter of academic debate. Measurement and curriculum inversion has been demonstrated repeatedly and consistently across the institutions responsible for HTA teaching, research, publication, and policy.

Recognition of this fact does not imply that change must occur overnight. Nor does it require the immediate abandonment of existing educational programs. What is required is a managed transition in which measurement principles are progressively introduced into HTA curricula and professional training. The starting point is straightforward: students should first understand the concept of an attribute and the distinction between manifest and latent attributes. This can be followed by an examination of the conventional scales of measurement and the conditions under which arithmetic operations are admissible.

Only then is it necessary to introduce representational measurement as the framework that explains why these requirements matter. Rather than presenting it as an abstract mathematical theory, it should be positioned as a critical tool for evaluating existing HTA practices. Students can then be encouraged to ask simple but fundamental questions. Why does multiplication require ratio measurement? Can a scale with negative values be a ratio scale? What properties must exist before a QALY can be constructed? How can latent attributes be measured? These questions do not require students to abandon current HTA methods. They require them to evaluate those methods critically.

The objective is therefore not disruption but reconstruction. By introducing measurement concepts sequentially and embedding critical appraisal throughout the curriculum, Canadian HTA can move from an environment where arithmetic precedes measurement to one where quantitative claims are grounded in the standards of normal science.

To facilitate this transition, Maimon Research has developed a concise nine-unit educational program designed specifically to support the movement from measurement inversion to

*measurement-based health technology assessment. The program is intended for faculty, researchers, HTA practitioners, graduate students, formulary committee members, and policy analysts who wish to understand the implications of representational measurement and its application to the assessment of therapeutic value*²⁰.

The objective is not to replace existing curricula overnight but to provide a structured pathway through which measurement concepts can be introduced sequentially and critically evaluated. Beginning with attributes and the conventional scales of measurement, the program progresses through representational measurement, dimensional homogeneity, manifest and latent attributes, ratio measurement, Rasch measurement, evaluable value claims, protocol design, and the requirements for replication and falsification. At each stage participants are encouraged to compare these principles with current HTA practices and assess the extent to which prevailing methodologies satisfy the requirements for quantitative claims.

The program therefore provides the foundation for a managed transition rather than a disruptive replacement of existing educational structures. Its purpose is to equip participants with the conceptual tools necessary to evaluate the scientific status of utilities, QALYs, simulation models, and other accepted HTA constructs. In doing so, it offers a practical framework through which Canadian HTA can move progressively toward standards that are consistent with representational measurement, normal science, and the generation of evaluable and replicable evidence.

CANADA: CONCLUSIONS

The findings presented in this assessment point to a clear conclusion. The challenge facing Canadian health technology assessment is not primarily one of refining existing methods, improving simulation models, or developing new utility instruments. The challenge is more fundamental. Across agencies, research centers, academic programs, professional journals, and training curricula, the concepts required to support valid quantitative claims are largely absent. The result is a pattern of measurement inversion in which arithmetic is routinely undertaken without first establishing the measurement properties required to support it.

Recognizing this problem is important, but it should not be viewed as a crisis. Nor does it imply that decades of HTA practice must be discarded overnight. Canadian HTA has accumulated considerable expertise in evidence evaluation, clinical assessment, policy analysis, and decision support. The task is not to abandon those achievements but to place them on a more secure scientific foundation.

The implication is that the future of HTA lies in a managed transition to measurement-based evaluation. This begins with introducing a clearer understanding of attributes, scale types, dimensionality, and the distinction between manifest and latent constructs. It requires greater attention to the principles of representational measurement and a willingness to examine critically the assumptions that underpin utilities, QALYs, and simulation models. Most importantly, it requires that measurement once again precede arithmetic.

The good news is that this transition is both practical and achievable. The principals involved are well established. The educational resources required already exist. The process can be

implemented progressively through professional development, curriculum enhancement, methodological training, and the gradual introduction of measurement-based standards into HTA practice. Institutions, researchers, educators, and decision-makers can move at a measured pace while maintaining continuity in teaching and policy activities.

The evidence presented here suggests that the current framework cannot be sustained indefinitely because it lacks the measurement foundations required for quantitative science. At the same time, there is no need for disruption or confrontation. The path forward is one of orderly reconstruction: building on existing strengths while introducing the principles necessary to support evaluable, replicable, and scientifically credible claims regarding therapeutic value. The goal is not simply to change HTA, but to strengthen it by aligning it with the standards of measurement that govern every other quantitative discipline.

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

² Krantz D, Luce R, Suppes P, Tversky A. Foundations of Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

³ The Theory of Allowable Transformations and the QALY. Logit Working Paper No 887. <https://maimonresearch.com/logit-working-paper-no-887-june-2026/>

⁴ Canada: National Endorsement of Measurement Failure in Health Technology Assessment. Logit Working Paper No 55. <https://maimonresearch.com/logit-working-paper-no-55/>

⁵ Canada: CADTH/CDA and the endorsement of measurement failure. Logit Working Paper No 57. <https://maimonresearch.com/logit-working-paper-no-57/>

⁶ Canada: Deconstructing the Epistemic Knowledge Base of the *Canadian Journal of Health Technologies (CJHT)*, Logit Working Paper No 2050. <https://maimonresearch.com/logit-working-paper-no-2050-february-2026/>

⁷ Canada: National endorsement of measurement failure in health technology assessment. Logit Working Paper No 55. <https://maimonresearch.com/logit-working-paper-no-55/>

⁸ Canada: The absence of representational measurement and the Toronto Health Economics and Technology Assessment Collaborative (THETA). Logit Working Paper No 8030. <https://maimonresearch.com/logit-working-paper-no-830-april-2026/>

-
- ⁹ Canada: The absence of representational measurement and the University Of Alberta Health Technology and Policy Unit (HTPU). Logit Working Paper No 834. <https://maimonresearch.com/logit-working-paper-no-834-april-2026/>
- ¹⁰ Canada: The absence of representational measurement and the Institute of Health Economics (IHE). Logit Working Paper No 833. <https://maimonresearch.com/logit-working-paper-no-833-april-2026/>
- ¹¹ Canada: The absence of representational measurement and the Dalhousie University Department of Community Health. Logit Working Paper No 835. <https://maimonresearch.com/logit-working-paper-no-835-april-2026/>
- ¹² Canada: The absence of representational measurement and the Centre for Clinical Epidemiology and Evaluation (C2E2) University of British Columbia. Logit Working Paper No 836. <https://maimonresearch.com/logit-working-paper-no-836-april-2026/>
- ¹³ Canada: The absence of representational measurement and the Institute for Clinical Evaluative Sciences (ICES). Logit Working Paper No 831. <https://maimonresearch.com/logit-working-paper-no-831-april-2026/>
- ¹⁴ Canada: Knowledge base consensus on measurement inversion in health technology assessment for six Canadian research centres. Logit Working Paper No 850. <https://maimonresearch.com/logit-working-paper-no-850-april-2026/>
- ¹⁵ Canada: Deconstructing the epistemic knowledge base of HUI Instruments. Logit Working Paper No 60. <https://maimonresearch.com/logit-working-paper-no-60/>
- ¹⁶Canada: HTAi and the Institutionalization of Curriculum Inversion. Logit Working Paper No 610. <https://maimonresearch.com/logit-working-paper-no-1610-june-2026/>
- ¹⁶ Canada: Health Technology Assessment International (HTAi) and the Commitment to False Measurement. Logit Working Paper No 59. <https://maimonresearch.com/logit-working-paper-no-59-february-2026/>
- ¹⁷ Canada: Deconstructing the epistemic knowledge base of the Canadian Journal of Health Technologies (CJHT). Logit Working Paper No 2050. <https://maimonresearch.com/logit-working-paper-no-2050-february-2026/>
- ¹⁸ Canada: Curriculum inversion as the foundation of measurement inversion. Logit Working Paper No 1255. <https://maimonresearch.com/logit-working-paper-no-1255-june-2026/>
- ¹⁹ Australia: There are only two lawful measures for therapy impact assessment. Logit Working Paper No 1811. <https://maimonresearch.com/logit-working-paper-no-1811-june-2026/>
- ²⁰ Mamon Research LLC. <https://maimonresearch.com/hta-reconstruction-program-and-fees/>