MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

NEW ZEALAND: PARADIGM FAILURE IN HEALTH TECHNOLOGY ASSESSMENT

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

LOGIT WORKING PAPER No 543 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper argues that health technology assessment (HTA) in New Zealand has reached a point of paradigm failure. The claim is based on evidence from a series of large language model interrogations undertaken at three levels: the national HTA environment, PHARMAC, and the principal university-based HTA research and training centers at the University of Auckland and the University of Otago. The results indicate a consistent pattern of measurement inversion and curriculum inversion that undermines the scientific foundations of the contemporary reference-case framework.

Measurement inversion occurs when arithmetic precedes measurement. Rather than establishing the measurement properties of a quantity before determining which mathematical operations are admissible, the reference-case framework assumes that measurement has already been achieved. Utilities are treated as though they possess ratio properties, QALYs are constructed through multiplication, and simulation models extend these assumptions across hypothetical populations and lifetime horizons. Yet the measurement requirements necessary to support these operations are not demonstrated. The consequence is a framework built upon numerical constructions rather than validated measures.

The curriculum interrogations reveal a parallel process of curriculum inversion. While students are introduced to utilities, QALYs, cost-effectiveness analysis and simulation modelling, there is little recognition of scales of measurement, representational measurement, unidimensionality, dimensional homogeneity, latent attribute measurement or ratio measurement. The concepts required to evaluate the scientific legitimacy of the reference case are therefore largely absent from the educational environment. Graduates are trained to apply the paradigm rather than assess its foundations.

Particular attention is given to the role of ratio measurement. The paper argues that the failure to recognize the constraints imposed by ratio scales is the critical error underpinning utility-based assessment. Once ratio measurement is restored to its central role in scientific inquiry, the assumptions supporting utilities, QALYs and reference-case simulation models can no longer be sustained. The resulting deficiencies are not methodological weaknesses that can be corrected through refinement. They are foundational defects that challenge the scientific legitimacy of the paradigm itself.

The paper concludes that the status quo in New Zealand HTA is unsustainable. The issue is not whether the reference case can be improved, but whether a framework founded upon arithmetic before measurement can survive once the standards of measurement science are restored. The only viable response is reconstruction. A new curriculum and analytical framework are required, grounded in representational measurement, valid ratio measures, latent and manifest attribute measurement, and evaluable and falsifiable claims regarding therapy impact.

To address this gap, Maimon Research LLC has developed a nine-unit educational program designed specifically to support curriculum reconstruction [1]*. The program introduces the*

principles of measurement science, distinguishes manifest from latent attributes, outlines the requirements for valid ratio measurement, presents Rasch measurement for latent attributes and provides a framework for developing evaluable and falsifiable claims regarding therapy impact. The objective is not to modify the reference case but to provide the educational foundation for its replacement.

INTRODUCTION

Health technology assessment (HTA) in New Zealand has reached a critical point. For more than four decades, HTA internationally has been dominated by a common analytical framework built around utility scores, quality-adjusted life years (QALYs), cost-effectiveness ratios and reference-case simulation models. In New Zealand, this framework is reflected in the policy environment associated with PHARMAC and in the academic and research settings that support HTA training, analysis and professional development . These constructs have achieved widespread institutional acceptance and now occupy a central role in reimbursement decisions, formulary assessments, academic research and professional education.

Institutional acceptance, however, is not the same as scientific validity. The central question confronting New Zealand HTA is no longer whether these methods are widely used, but whether they satisfy the standards required for quantitative claims. This distinction is crucial because quantitative science begins with measurement. Before arithmetic operations can be undertaken, it must first be demonstrated that the quantities entering those operations possess the properties necessary to support them. Measurement precedes arithmetic.

The argument advanced in this paper is straightforward. New Zealand HTA, like the wider international HTA community, has developed around a systematic inversion of this principle. Rather than establishing measurement and deriving admissible forms of arithmetic from demonstrated scale properties, the field has proceeded in the opposite direction. Arithmetic operations have been accepted first, while the measurement status of the underlying quantities has largely been assumed. Utilities are treated as though they possess ratio properties. QALYs are treated as though multiplication is admissible. Reference-case simulation models are treated as though they generate quantitative evidence. Yet the measurement foundations necessary to support these claims have never been demonstrated.

This failure has consequences that extend far beyond individual analytical techniques. Once measurement is displaced by arithmetic, every subsequent stage of the analytical process inherits the same defect. Utility weights, QALYs, incremental cost-effectiveness ratios and simulation-based projections become part of a chain of dependent assumptions rather than a chain of validated measures. The resulting framework may generate numbers, but the generation of numbers is not equivalent to measurement.

The purpose of this paper is to argue that HTA in New Zealand has reached the end point of this process. The issue is no longer one of methodological refinement, improved modelling techniques or alternative utility instruments. The problem is foundational. The assumptions that support the contemporary reference-case paradigm are incompatible with the standards of representational measurement. Once those standards are restored, the conclusion becomes unavoidable: the present paradigm cannot be sustained.

Paradigm failure does not imply the disappearance of PHARMAC, university research centers, journals or professional communities. These may continue to operate for many years. Paradigm failure occurs when the assumptions that support a field are shown to be incompatible with the standards of science. In the case of New Zealand HTA, the critical issue is not whether the reference case remains popular or institutionally entrenched. The issue is whether utilities, QALYs and simulation models satisfy the requirements of measurement. Once these requirements are examined, the conclusion is unavoidable. The contemporary HTA paradigm is sustained not by demonstrated measurement but by assumptions regarding measurement.

The challenge facing New Zealand HTA is not impossible reform but reconstruction. The reference case offers no basis for reform because the source of failure lies in its design. Utilities, QALYs and reference-case simulation models are not imperfect implementations of sound measurement principles. They are constructs built upon assumptions that fail the requirements of measurement. The problem is not methodological weakness but architectural failure.

Reconstruction therefore becomes imperative. It cannot be avoided if HTA is to retain any credibility. 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 is a 40-year design failure. After more than four decades, the absence of measurement can no longer be treated as an oversight, a methodological limitation or a topic for future discussion. The evidence demonstrates that the paradigm itself is incapable of satisfying the standards required for quantitative science.

NEW ZEALAND: DEFINING PARADIGM FAILURE

A paradigm fails when the assumptions that support it can no longer be reconciled with the standards that govern scientific inquiry. The failure may not be immediately visible. Institutions may continue to employ the paradigm, journals may continue to publish within it, and educational programs may continue to teach it. Yet once its foundational assumptions are shown to be inconsistent with accepted scientific standards, the paradigm loses its intellectual legitimacy. It was always a design failure,

Paradigm failure is therefore not a matter of disagreement over methods or competing analytical preferences. It 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.

In the physical sciences, paradigm failure occurs when observations repeatedly conflict with a theoretical structure and the resulting anomalies cannot be accommodated within the existing framework. In measurement science, paradigm failure occurs when the quantities upon which a framework depends are shown not to possess the properties required for the operations being performed. The problem is not that the calculations are inaccurate. The problem is that the calculations are inadmissible.

Applied to health technology assessment, paradigm failure occurs when the central constructs of the field are shown to violate the requirements of measurement scales and representational measurement [2] [3] [4] . Utilities are treated as though they possess ratio properties without demonstration. QALYs are constructed through multiplication without establishing that multiplication is admissible. Reference-case simulation models extend these assumptions across hypothetical lifetimes and populations. If the measurement requirements necessary to support these operations are absent, then the framework loses its scientific foundation.

The relevance to New Zealand is immediate. PHARMAC employs a reference-case framework in which utility-based evaluations and QALY-derived assessments play a central role in reimbursement decisions [5]. The practice interrogations reported in this paper for research centers indicate that the educational and research environments supporting HTA in New Zealand similarly give little recognition to representational measurement, scale theory, unidimensionality, latent attribute measurement and ratio measurement [6] [7] [8] . The 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; in short, historical curiosities.

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 New Zealand, this conclusion applies equally to the reference-case framework employed by PHARMAC and to the educational and research structures that continue to reproduce it. The issue is not whether the current paradigm can be improved. The issue is whether a framework constructed on arithmetic before measurement can survive once measurement is restored to its central place in scientific inquiry. If the answer is no, then what is being observed is not methodological weakness but paradigm failure.

The implications are unavoidable. The future of HTA in New Zealand 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 [9].

NEW ZEALAND: THE EVIDENCE FOR MEASUREMENT INVERSION

The evidence for measurement inversion in New Zealand is both extensive and consistent. It is not confined to a single institution, a particular methodological preference or an isolated policy framework. Rather, it is evident across the entire HTA environment. Interrogations undertaken at the national level, together with separate assessments of PHARMAC, the University of Auckland and the University of Otago, reveal a common pattern: quantitative claims are accepted and employed without first demonstrating that the quantities involved satisfy the requirements of measurement.

As detailed in the individual interrogation studies, the extent of measurement inversion was assessed through responses to a set of 24 canonical statements drawn from the principles of representational measurement and the requirements for valid quantitative claims. These statements comprised both true and false propositions. The true propositions reflected accepted principles of measurement science, while the false propositions represented assumptions embedded within the contemporary reference-case paradigm that are inconsistent with those principles.

The logic of the assessment was straightforward. In a knowledge base aligned with the standards of measurement science, the true propositions should attract high levels of endorsement and the false propositions should attract low levels of endorsement. Such a pattern would indicate recognition of the requirements for measurement, admissible arithmetic and valid quantitative claims.

The interrogations revealed the opposite pattern. Statements that should have been strongly endorsed received consistently low endorsement probabilities, while statements that should have been rejected received consistently high endorsement probabilities. The result was not merely a lack of knowledge regarding measurement theory. It was evidence of reversal. The knowledge bases systematically favored propositions that conflicted with the principles of representational measurement while failing to endorse propositions that reflected those principles.

This reversal is the defining characteristic of measurement inversion. The issue is not simply that measurement concepts are absent. The issue is that the analytical framework actively supports assumptions that contradict the requirements of measurement science. The interrogations therefore provide evidence not only of omission but of endorsement of error. Knowledge bases associated with HTA agencies, research centers and educational programs repeatedly accepted propositions that treated arithmetic as independent of measurement, assumed ratio properties where none had been demonstrated and supported quantitative claims without first establishing the measurement status of the underlying quantities.

The significance of this finding is considerable. If the interrogations had merely shown low awareness of measurement principles, the conclusion might have been limited to deficiencies in curriculum coverage or professional training. Instead, the results indicate the presence of an intellectual framework that systematically privileges false measurement propositions over true ones. The problem is therefore not ignorance alone. It is the institutionalization of assumptions that are incompatible with the standards of quantitative science that has lasted for some 40 years.

The evidence for measurement inversion rests on this reversal. True propositions are rejected. False propositions are endorsed. The resulting pattern demonstrates that the reference-case paradigm is not simply disconnected from measurement science; it operates according to assumptions that are fundamentally opposed to it.

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 New Zealand 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.

The separate interrogation of PHARMAC reached a similar conclusion. PHARMAC's decision-making framework relies heavily upon utility-based evaluations and QALY-derived assessments. These constructs occupy a central role in determining comparative value and informing reimbursement decisions. Yet their use presupposes that utility scores possess ratio properties capable of supporting multiplication by time. No demonstration of these properties is provided. The legitimacy of the arithmetic is assumed rather than established.

This assumption is critical because the QALY depends entirely upon multiplication. Survival time is unquestionably a ratio measure. The utility score is therefore required to function as a proportional adjustment factor. Unless the utility score itself possesses ratio properties, the multiplication is inadmissible. The resulting QALY becomes a numerical construction rather than a measure. The PHARMAC interrogation revealed no evidence that this foundational issue occupies a meaningful place within the analytical framework.

The interrogations of the University of Auckland and the University of Otago point to the same conclusion. Both institutions support research and training activities associated with HTA. Both contribute to the intellectual environment from which future researchers, analysts and policy advisers emerge. Yet neither interrogation revealed meaningful recognition of the measurement requirements necessary to support utility-based claims. Utilities, QALYs and simulation modelling are accepted components of analysis, while the measurement foundations upon which they depend remain largely invisible.

The significance of this consistency should not be underestimated. Independent interrogations of policy agencies, research centers and academic institutions all point in the same direction. The issue is not confined to one organization or one group of researchers. It reflects a broader intellectual framework in which arithmetic has displaced measurement as the starting point for quantitative claims.

The consequences become particularly apparent when examining the role of utilities. Utility scores are routinely interpreted as though they represent proportions of health. A utility value of 0.8 is interpreted as 80 percent of full health; a value of 0.5 as 50 percent of full health. Yet proportional interpretation requires ratio measurement. A ratio measure requires a fixed non-arbitrary zero and meaningful proportional comparisons. Neither condition has been demonstrated for utility scores. The proportional interpretation is simply assumed and then embedded within the arithmetic of the QALY. The result is a number not a measure.

The same problem extends to reference-case simulation models. These models do not generate new measures. They manipulate quantities already assumed to be valid. If the utility scores entering the model lack ratio properties, every QALY generated by the model inherits the same defect. Increasing analytical sophistication cannot compensate for defective inputs. A simulation model cannot create measurement where measurement does not exist.

What makes the New Zealand evidence particularly compelling is its uniformity. Whether one examines HTA nationally, PHARMAC specifically or the two principal university-based research environments, the same pattern emerges. There is no identifiable point at which the requirements of representational measurement are systematically asserted. There is no identifiable point at which utility scores are required to demonstrate ratio properties. There is no identifiable point at which the admissibility of QALY arithmetic is established.

This is the defining characteristic of measurement inversion. Numerical operations are accepted first, while measurement is assumed to follow automatically. Arithmetic acquires authority simply because it is numerical. The distinction between a number and a measure disappears.

The consequence is that the central constructs of New Zealand HTA rest upon assumptions rather than demonstrated measurement. Utilities are assumed to be ratio measures. QALYs are assumed to be valid quantitative outcomes. Simulation models are assumed to generate evidence. Yet the measurement requirements necessary to support these assumptions remain unmet.

Taken together, the national interrogation and the separate assessments of PHARMAC, Auckland and Otago provide compelling evidence that measurement inversion is not an isolated methodological weakness but a defining characteristic of the New Zealand HTA paradigm. The problem is therefore not one of implementation or technical refinement. It is foundational. The paradigm rests upon arithmetic before measurement. Once measurement is restored to its central role in scientific inquiry, the deficiencies become impossible to ignore. Measurement inversion is not merely present within New Zealand HTA; it is embedded within the framework itself. This is why the issue is no longer one of reform but of paradigm failure.

NEW ZEALAND: THE EVIDENCE FOR CURRICULUM INVERSION

If measurement inversion explains why contemporary HTA fails the standards of measurement science, curriculum inversion explains why that failure has persisted. The two phenomena are closely connected. Measurement inversion concerns the analytical framework itself. Curriculum inversion concerns the educational environment that reproduces and sustains that framework. Together they provide compelling evidence of paradigm failure in New Zealand HTA [10].

The curriculum-related knowledge bases associated with PHARMAC, the University of Auckland and the University of Otago were interrogated using a set of ten canonical curriculum statements. These statements were designed to represent the minimum concepts that should be present in any curriculum intended to support scientifically credible claims regarding therapy impact. The statements addressed attributes, scales of measurement, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement, Rasch measurement and falsifiable claims. The objective was not to assess individual courses or instructors but to determine whether these foundational concepts were embedded within the broader curriculum knowledge base that supports teaching, research and professional development in HTA.

The logic of the assessment was straightforward. A curriculum aligned with the standards of measurement science would be expected to show strong endorsement of concepts central to measurement and quantitative reasoning. High endorsement probabilities would indicate that students and researchers are likely to encounter these principles as part of their professional training. Conversely, low endorsement probabilities would indicate that the concepts are largely absent from the curriculum environment. The interrogation results revealed a consistent pattern across PHARMAC, Auckland and Otago. While there was recognition of attributes and, to a lesser extent, falsifiable claims, there was little recognition of representational measurement, scale theory, unidimensionality, latent attribute measurement, ratio measurement and the distinction between manifest and latent forms of measurement. The findings point to curriculum inversion, where the concepts required to evaluate the scientific legitimacy of the reference-case paradigm are largely absent from the educational structures responsible for reproducing that paradigm.

A scientific curriculum should equip students with the concepts necessary to evaluate the validity of quantitative claims. In HTA, this would require instruction in attributes, scales of measurement, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement, dimensional homogeneity and falsifiable claims. These concepts are not optional additions to professional training. They are the foundations upon which all quantitative reasoning depends. Without them, students may learn analytical techniques, but they cannot determine whether the outputs generated by those techniques are scientifically meaningful.

The curriculum interrogations undertaken for PHARMAC, the University of Auckland and the University of Otago indicate that these foundations are largely absent. While there is moderate recognition of outcome identification and target-attribute specification, the principles necessary to establish whether quantitative claims are possible receive little endorsement. The curriculum therefore introduces students to the practice of HTA without first introducing them to the science of measurement.

This omission is evident in the treatment of measurement scales. Across all three interrogations, there is little indication that students are systematically exposed to the differences between nominal, ordinal, interval and ratio scales or to the implications these distinctions have for quantitative analysis. Yet these distinctions determine whether arithmetic operations are admissible. Without them, there is no basis for deciding whether multiplication, division or averaging can be performed meaningfully. The result is that students are likely to encounter numerical outputs without understanding the scale properties upon which those outputs depend.

The same pattern is evident in the treatment of representational measurement. The proposition that quantitative claims must be grounded in the axioms of measurement receives almost no endorsement. This finding is particularly important because representational measurement provides the scientific framework that distinguishes measures from numerical assignments. It establishes the conditions under which arithmetic is lawful and determines whether quantitative claims are defensible. Its absence leaves students without the conceptual tools necessary to evaluate the legitimacy of the methods they are taught to apply.

The interrogation results also reveal little recognition of unidimensionality. Measurement requires that an attribute represent a single dimension. Without unidimensionality, numerical aggregation may be possible, but measurement is not. Yet students are exposed to multidimensional instruments, composite indices and utility-based frameworks without systematic instruction in this requirement. The consequence is that aggregation is easily mistaken for measurement.

Equally significant is the treatment of manifest and latent attributes. The curriculum gives only limited recognition to the distinction between directly observable outcomes and latent constructs that require a measurement model. More importantly, there is almost no recognition that manifest and latent attributes require fundamentally different measurement frameworks. . Students are therefore unlikely to appreciate that survival, hospital admissions and treatment persistence pose different measurement problems from pain, fatigue, quality of life or patient functioning. The distinction between direct observation and latent attribute measurement is largely lost.

These omissions are not random. They occur precisely in those areas most likely to challenge the assumptions that underpin the reference-case paradigm. Students are taught utilities, QALYs, cost-effectiveness analysis and simulation modelling. They are not taught the measurement principles necessary to determine whether utilities support proportional interpretation, whether QALYs satisfy dimensional homogeneity or whether simulation outputs are grounded in valid measures. The result is an educational framework that reproduces the assumptions of the reference case while neglecting the concepts required to evaluate it critically.

This is the essence of curriculum inversion. Instead of beginning with measurement and proceeding to analysis, the curriculum begins with analytical techniques and assumes that measurement has already been established. The educational sequence mirrors the analytical sequence observed in measurement inversion. Arithmetic precedes measurement in practice because arithmetic precedes measurement in education.

The consequences extend beyond the classroom. Graduates move into research centers, policy agencies and professional organizations carrying the same conceptual limitations. They become researchers, reviewers, educators and decision makers. The assumptions embedded in the curriculum are then reproduced through research publications, HTA submissions, reimbursement decisions and future educational programs. Curriculum inversion therefore becomes the mechanism through which measurement inversion is sustained from one generation to the next.

The consistency of the New Zealand findings is particularly revealing. PHARMAC, Auckland and Otago display remarkably similar profiles. Although differing in institutional role and function, all three show strong recognition of outcomes and evaluation while giving little attention to measurement science. This uniformity suggests that the problem is systemic rather than local. It reflects the structure of the HTA paradigm itself rather than the practices of any individual institution.

The evidence therefore points to a clear conclusion. New Zealand HTA does not merely suffer from measurement inversion; it reproduces that inversion through its educational and research environment. Students are trained to operate within the reference-case framework without being equipped to evaluate its measurement foundations. The result is a self-sustaining cycle in which analytical methods are taught, applied and defended while the scientific principles necessary to assess their validity remain largely absent.

Curriculum inversion thus provides the missing explanation for the persistence of measurement inversion. The former reproduces the latter. Together they reveal a paradigm that has become detached from the standards of measurement science while continuing to present itself as a quantitative discipline. This is why the challenge facing HTA in New Zealand is not merely methodological reform. It is the reconstruction of both the analytical framework and the curriculum that sustains it.

NEW ZEALAND: THE ABSENCE OF RASCH MEASUREMENT

One of the most striking findings from the New Zealand measurement and curriculum interrogation is the the near-complete absence of Rasch ratio measurement and its role in the assessment of latent attributes [11]. This omission is 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.

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 interrogation suggests that this perspective is absent from the New Zealand 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 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 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.

NEW ZEALAND: UNDERSTANDING RATIO MEASUREMENT

The fatal weakness of contemporary health technology assessment (HTA) in New Zealand is not a failure of modelling, statistical analysis or computational sophistication. It is a failure to understand ratio measurement. 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 New Zealand 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 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.

The failure to understand the central importance of ratio measurement is compounded by an equally serious failure: the absence of Rasch measurement from the conceptual foundations of health technology assessment. Rasch measurement provides the necessary and sufficient conditions for transforming observations or ordinal responses into measures capable of supporting quantitative claims regarding latent attributes. Its objective is the construction of invariant measures of attribute possession through a rigorous measurement model.

This omission is particularly striking because many of the outcomes that HTA claims to value are inherently latent. Pain, fatigue, physical functioning, emotional well-being, quality of life and need fulfilment are not directly observable attributes. They cannot be measured through simple counting or direct observation. They require a measurement framework capable of estimating possession of the attribute while satisfying the requirements of unidimensionality, invariance and measurement. Rasch measurement was developed precisely for this purpose.

Over the past six decades, Rasch methods have become established across education, psychology, rehabilitation, health outcomes research and many other fields where latent attributes are central to evaluation. Yet despite its relevance to patient-centered outcomes, Rasch measurement remains absent from mainstream HTA. Instead, the field has relied upon utility instruments, preference scores and composite indices that fail to satisfy the requirements of measurement while simultaneously ignoring the only framework specifically designed to create measures of latent attribute possession.

The implications are profound. HTA has attempted to quantify subjective outcomes without adopting the measurement science required to support those claims. The result is a discipline that routinely employs ordinal preference scores as though they were measures while neglecting the measurement model capable of transforming observations into valid measures. This is not merely a methodological oversight. It represents a fundamental failure to appreciate the distinction between numerical scoring and measurement.

Given the central importance of subjective patient outcomes in contemporary HTA, the absence of Rasch measurement from the field is difficult to justify. Once the requirements of ratio measurement are recognized, the omission becomes even more serious. HTA has simultaneously ignored the conditions required for ratio measurement of manifest attributes and the conditions required for ratio measurement of latent attributes. The consequence is that the field lacks a scientifically defensible measurement framework for either category of outcome. This omission lies at the heart of paradigm failure and reinforces the case for reconstruction rather than reform.

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 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 New Zealand is not reform but reconstruction. Ratio measurement does not merely challenge the reference case; it exposes why the reference case cannot survive.

NEW ZEALAND: 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. New Zealand is no exception with the examples of PHARMAC and the Universities of Auckland and Otago. The belief in the QALY is still very much alive.

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 memeplex; a complex of ideas, behaviors, or cultural practices that spread together. 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 global 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.

NEW ZEALAND: 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 New Zealand 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.

The New Zealand curriculum interrogations indicate that these tools are largely absent. The concepts most essential for evaluating the reference case receive the weakest endorsement. Scales of measurement are not central. Representational measurement is effectively invisible. Unidimensionality receives minimal recognition. Latent attribute measurement is poorly represented. The distinction between manifest and latent forms of ratio measurement is almost absent. As a result, the curriculum does not prepare graduates to evaluate the measurement foundations of HTA. It prepares them to apply the reference case; a professional career locked inside the memeplex,

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 further consequence of paradigm failure is that the educational resources required for reconstruction are largely absent. Existing HTA textbooks and teaching materials were developed within the reference-case tradition and therefore reproduce many of the assumptions that have contributed to measurement inversion. They provide extensive coverage of utilities, QALYs, cost-effectiveness analysis and simulation modelling, but virtually no coverage of representational measurement, ratio measurement, dimensional homogeneity, Rasch measurement, ,manifest and latent attributes or falsifiable claims. As a consequence, there is currently no widely adopted textbook framework available to guide the transition from arithmetic-based assessment to measurement-based assessment. *To address this gap, Maimon Research LLC has developed a nine-unit educational program designed specifically to support curriculum reconstruction* [12]*. The program introduces the principles of measurement science, distinguishes manifest from latent attributes, outlines the requirements for valid ratio measurement, presents Rasch measurement for latent attributes and provides a framework for developing evaluable and falsifiable claims regarding therapy impact. The objective is not to modify the reference case but to provide the educational foundation for its replacement.*

For New Zealand, the implications are direct. PHARMAC, Auckland and Otago occupy important positions in policy, research and professional development. If these institutions continue to reproduce a curriculum that omits measurement science while endorsing reference-case methods, the result will be continued measurement inversion. Graduates will enter professional roles able to conduct analyses but unable to judge whether the analyses rest on valid measures.

NEW ZEALAND: CONCLUSION

The evidence presented in this report leads to a conclusion that is both straightforward and unavoidable. The *status quo* in New Zealand health technology assessment is unsustainable. This is not because HTA lacks technical sophistication, computational power or institutional support. Nor is it because individual analysts, researchers or policy makers have failed to apply accepted methods correctly. The problem lies deeper. The analytical foundations of the contemporary reference-case framework are incompatible with the requirements of measurement science.

The interrogations undertaken at the national level, together with the separate assessments of PHARMAC, the University of Auckland and the University of Otago, reveal a consistent pattern of measurement inversion and curriculum inversion. Arithmetic is accepted before measurement. Utilities are treated as though they possess ratio properties without demonstration. QALYs are constructed through multiplication without establishing that multiplication is admissible. Reference-case simulation models extend these assumptions across hypothetical populations and lifetime horizons. At the same time, the educational environment provides little exposure to the scales of measurement, representational measurement, unidimensionality, dimensional homogeneity and ratio measurement that would allow these assumptions to be critically examined.

The result is a self-reinforcing paradigm, a global memeplex, in which the reference case is simultaneously taught, applied and defended while the scientific principles necessary to evaluate its legitimacy remain largely absent. Students are trained to operate within the framework but not to assess its foundations. Researchers refine methods whose measurement status is assumed rather than demonstrated. Policy agencies rely upon numerical constructions that are treated as evidence despite the absence of valid measures.

This position cannot be sustained indefinitely. Once the requirements of measurement are restored to their central place in scientific inquiry, the weaknesses of the reference case become impossible to ignore. The problem is not that particular assumptions require adjustment or that individual methods require refinement. The problem is that the framework itself rests upon quantities whose measurement properties have never been established. Arithmetic cannot create measurement where measurement does not exist.

The implications for New Zealand are therefore clear. The future of HTA does not lie in preserving the reference case through increasingly sophisticated modelling exercises. Nor does it lie in incremental methodological reform. The necessary response is reconstruction. Professional education, research and policy must be rebuilt around the principle that measurement precedes arithmetic. Quantitative claims must be grounded in valid measures, whether through direct ratio measurement of manifest attributes or appropriate measurement models for latent attributes. Claims must be evaluable, replicable and capable of falsification.

The choice is now unavoidable. New Zealand HTA can continue to preserve the reference case by ignoring the requirements of measurement, or it can preserve the standards of science by abandoning the reference case. It cannot do both. The *status quo* has reached its end point. The era of measurement inversion must give way to a framework built upon measurement itself.

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] Mamon Research LLC. https://maimonresearch.com/hta-reconstruction-program-and-fees/

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

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

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

[5] New Zealand: PHARMAC endorses false measurement for health technology assessment. Logit Working Paper No 1261. https://maimonresearch.com/logit-working-paper-no-1261-april-2026/

[6] New Zealand. National endorsement of measurement absence in health technology assessment. Logit Working Paper No 35. file:///C:/Users/Paul/Downloads/Logit-Working-Paper-No-35-January-2026-REV.pdf

[7] New Zealand: The absence of representational measurement in health technology assessment at the University of Auckland. Logit Working Paper No 779. https://maimonresearch.com/logit-working-paper-no-779-may-2026/

[8] New Zealand. The absence of representational measurement in health technology assessment at the University of Otago. Logit Working Paper No 778. https://maimonresearch.com/logit-working-paper-no-778-may-2026/

[9] New Zealand. There are only two lawful measures for therapy impact assessment. Logit Working Paper No 1813. https://maimonresearch.com/logit-working-paper-no-1813-june-2026/

[10] New Zealand. The missing science of measurement in health technology assessment, Logit Working Paper No 1277. https://maimonresearch.com/logit-working-paper-no-1277-june-2026/

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

[12] Mamon Research LLC. https://maimonresearch.com/hta-reconstruction-program-and-fees/