MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

AUSTRALIA: THE PBAC AND AUSTRALIAN HTA RESEARCH CENTERS – THE INVISIBILITY OF RASCH MEASUREMENT

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

LOGIT WORKING PAPER No 179 MAY 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper examines a neglected question in the history of health technology assessment (HTA): why has Rasch measurement remained effectively invisible within a discipline heavily dependent upon patient-reported outcomes, quality-of-life assessment and other latent attributes? For more than sixty years the Rasch framework has provided a scientifically rigorous method for transforming subjective observations into invariant quantitative measures. Widely applied in education, psychology and rehabilitation medicine, Rasch measurement offers the only established framework for constructing lawful measures of latent attributes consistent with the requirements of representational measurement. Yet despite the centrality of subjective outcomes to HTA, Rasch concepts remain largely absent from its teaching, research and policy frameworks.

To investigate this paradox, the paper employs large language model interrogation of HTA knowledge bases. Five canonical statements capturing the essential features of Rasch measurement were applied to the UK institutions that shaped modern HTA—the Office of Health Economics (OHE), the York Centre for Health Economics (CHE) and the National Institute for Health and Care Excellence (NICE)—together with the Australian Pharmaceutical Benefits Advisory Committee (PBAC) and four leading Australian HTA research centers. Endorsement probabilities were used to assess the degree to which Rasch concepts are recognized within these knowledge structures.

The findings reveal a remarkable pattern of uniformity. Core Rasch propositions receive endorsement probabilities at or near floor values across all institutions examined. While there is limited recognition that latent attributes exist, there is almost no recognition of Rasch measurement as the framework required to evaluate them. The results suggest that Rasch measurement was never incorporated into the conceptual foundations of HTA. Instead, HTA evolved around valuation of health states, utility construction, QALYs and reference-case simulation modelling.

The paper argues that the invisibility of Rasch has had profound consequences. By neglecting the measurement of latent attributes, HTA institutionalized a framework characterized by measurement inversion and arithmetic chaos, where arithmetic operations routinely precede lawful measurement. The proposed solution is straightforward: abandon the reference-case architecture and return HTA to measurable attributes, with manifest claims supported by linear ratio measures and latent claims supported by Rasch logit ratio measures. Rasch is therefore presented not merely as an alternative methodology, but as the exemplar of a reconstructed HTA grounded in measurement, falsification and the continuing evolution of objective knowledge.

INTRODUCTION

The history of modern health technology assessment (HTA) presents a curious puzzle. For more than sixty years the Rasch model has been recognized as the necessary and sufficient set of rules for transforming subjective observations into measures. Developed originally by Georg Rasch and subsequently refined through extensive theoretical and practical application, the Rasch framework provides a rigorous approach to constructing invariant measurement scales for latent attributes. It has been applied widely in education, psychology and rehabilitation medicine, Yet despite the central importance of subjective outcomes to HTA, Rasch measurement remains effectively invisible within the discipline.

This invisibility is difficult to explain. HTA places considerable emphasis on patient-reported outcomes, health-related quality of life, symptom burden, treatment satisfaction and other latent attributes. These are not directly observable phenomena but manifestations of underlying constructs. If such attributes are to support claims for therapy impact, they require lawful measurement. For six decades Rasch measurement has provided the only recognized framework capable of transforming ordinal subjective responses into invariant quantitative measures while satisfying the requirements of representational measurement. One might therefore expect Rasch concepts to occupy a central position within HTA teaching, research and policy development.

The evidence suggests precisely the opposite. Interrogations of the knowledge bases associated with the Pharmaceutical Benefits Advisory Committee (PBAC), leading Australian HTA research centers and the UK institutions that shaped modern HTA reveal minimal endorsement of core Rasch propositions. Statements concerning Rasch logit ratio measurement, latent trait possession and the transformation of subjective responses to measurement consistently receive floor-level endorsement probabilities. At the same time, utilities, QALYs and reference-case simulation models dominate the analytical framework. The question therefore is not whether Rasch has been rejected following careful consideration. Rather, the question is why Rasch appears never to have entered the conceptual architecture of HTA in the first place.

This paper examines that question. It argues that the invisibility of Rasch measurement in Australian HTA reflects inheritance of a UK-derived intellectual tradition centered on valuation rather than measurement. From the Office of Health Economics, through the York Centre for Health Economics and subsequently NICE, the dominant concern became valuation of health states, utility construction and cost-effectiveness modelling. The prior question—how latent attributes should be measured—was largely absent. Australian HTA inherited this framework. The consequence is a discipline in which Rasch measurement, despite its established role in transforming subjective observations into measures, remains largely outside the boundaries of accepted discourse.

THE GENESIS OF RASCH MEASUREMENT

The origins of Rasch measurement lie in a problem largely ignored by conventional measurement theory in the first half of the twentieth century: how can subjective observations be transformed into measures? By the 1940s the classification of measurement scales had been formalized through the work of. Stevens. His influential 1946 paper distinguished nominal, ordinal, interval and ratio scales and specified the arithmetic operations appropriate to each [i]. Stevens provided a powerful framework for understanding the properties of observed variables, but his account largely concerned manifest attributes where the object of measurement was directly observable. Little attention was given to latent attributes such as ability, symptom burden, quality of life or need fulfilment. The central question remained unresolved: how could one construct a measure for attributes that could not be observed directly?

This question became increasingly important as the social and behavioral sciences expanded. Researchers relied heavily upon questionnaires, ratings and patient reports. Numerical scores were generated by summing responses across multiple items, with the resulting totals frequently treated as though they represented quantitative measures. Yet there was no justification for assuming that a numerical score derived from subjective observations possessed interval or ratio properties. Adding responses together merely created another ordinal structure. The underlying problem therefore remained: how could subjective observations be transformed into a lawful measure?

It was this challenge that Georg Rasch addressed in his 1960 work *Probabilistic Models for Some Intelligence and Attainment Tests* [ii]. Rasch proposed a remarkably simple but powerful solution. Instead of assuming that numerical scores represented measurement, he specified a model that observations had to satisfy before measurement could be claimed. The model linked person ability and item difficulty within a single probabilistic structure. If the data fitted the model, then the resulting scale possessed invariant properties independent of the particular sample of respondents or items employed. Measurement was therefore not assumed; it had to be demonstrated.

The importance of this contribution is difficult to overstate. Rasch provided the first coherent framework for constructing measures of latent attributes while preserving the principles required for scientific inquiry. Subjective observations could now be transformed into a ruler with meaningful quantitative properties. The focus shifted from aggregate scores to the possession of an underlying attribute located on a latent continuum. Responses became evidence supporting measurement rather than measures in themselves.

The significance of Rasch's achievement became even clearer through the work of Benjamin Wright and colleagues at the University of Chicago during the 1960s and 1970s. Wright demonstrated that Rasch measurement was not merely a useful statistical model but a measurement framework possessing deep connections with the developing theory of representational measurement [iii]. As the axioms of representational measurement were formalized by Krantz, Luce, Suppes and Tversky in 1971, it became increasingly apparent that the requirements imposed by the Rasch model and the requirements imposed by representational measurement were formally equivalent [iv]. Both demanded unidimensionality, invariance, meaningful quantitative structure and the possibility of lawful arithmetic operations.

This insight was crucial. Rasch measurement solved the very problem that conventional scale theory left unresolved. Stevens had described the properties of measurement scales; Rasch provided a method for constructing them in the case of latent attributes. Representational measurement specified the conditions under which arithmetic operations were admissible; Rasch created instruments capable of satisfying those conditions. Together they established a coherent framework for latent measurement that remains unique.

The contrast with contemporary HTA could hardly be greater. Rather than constructing instruments capable of transforming subjective observations into lawful measures, HTA largely accepted ordinal scores, utility weights and preference values as though they already possessed the required quantitative properties. The challenge that motivated Rasch, how to transform subjective observations into measures, was effectively bypassed. Instead of measurement, HTA embraced valuation. Instead of latent attribute possession, it embraced preference aggregation. Instead of demonstrating measurement properties, it assumed them.

The consequence is that while Rasch measurement evolved into a mature scientific framework used internationally across multiple disciplines, HTA continued to rely upon ordinal and composite structures that fail the standards required for lawful measurement. The irony is striking. The problem that Rasch set out to solve more than sixty years ago is precisely the problem that remains unresolved within modern HTA.

INTERROGATING BELIEF IN RASCH MEASUREMENT

The central question addressed in this paper is straightforward: to what extent is Rasch measurement recognized within the intellectual architecture of HTA? This is not a question that can be answered simply by reviewing isolated publications, textbook references or methodological guidelines. The issue concerns the broader knowledge base that shapes teaching, research, policy recommendations and methodological development. The challenge is therefore to identify whether Rasch concepts are embedded within the conceptual structure of HTA or whether they remain largely absent from accepted discourse.

Recent developments in large language artificial intelligence provide a means of addressing this question. Large language models can be interrogated systematically to assess the extent to which specific propositions are reinforced, neglected or contradicted within a defined knowledge base. The objective is not to assess personal beliefs or institutional declarations. Rather, the objective is to determine how a target knowledge base behaves when confronted with statements drawn from a particular conceptual framework.

For the present analysis, a series of canonical statements was selected to capture the essential features of Rasch measurement. These statements address the core elements of the Rasch framework: the distinction between manifest and latent attributes, transformation of subjective responses into measurement, the Rasch logit ratio scale, possession of latent traits and the relationship between Rasch measurement and representational measurement. Each statement was then interrogated within the knowledge bases of selected HTA institutions.

The statements are:

- There are only two classes of measurement: linear ratio and Rasch logit ratio
- Transforming subjective responses to interval measurement is only possible with Rasch rules
- The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits
- The outcome of interest fir latent traits is the possession of that trait
- The Rach rules for measurement are identical to the axioms of representational measurement

The targets were chosen deliberately. The Office of Health Economics, London (OHE) [v], the Centre for Health Economics (CHE) at the University of York [vi] and the National Institute for Health and Care Excellence, London (NICE) [vii] represent the intellectual foundations of modern

HTA. Together they shaped the utility-based and reference-case framework subsequently adopted throughout much of the world. In effect, they constitute the first generation of HTA institutions. Australian HTA emerged largely within this intellectual tradition. The PBAC and major Australian HTA research centers therefore represent subsequent generations built upon the original UK framework.

For each statement, the interrogation process generates a categorical endorsement probability ranging from minimal endorsement to strong endorsement. These probabilities provide an indication of the extent to which the concepts embodied in the statement are reinforced within the target knowledge base. High probabilities indicate strong conceptual support; low probabilities indicate neglect, rejection or absence. When viewed collectively, the resulting endorsement profiles provide a picture of the visibility or invisibility of Rasch measurement within the HTA tradition. The following sections present these endorsement patterns for the UK institutions that established the HTA framework and for the Australian institutions that subsequently inherited it.

Including the PBAC[viii], the five Australian HTA research center are Melbourne University School of Population and Global Health (MSGPH) [ix], The Centre for Health Economics (CHE) Monash University [x], Health Economics Group (HEG) University of Adelaide [xi], Centre for Applied Health Economics (CAHE) Griffith University [xii] and the Leeder Centre for Health Policy, University of Sydney [xiii] .

ENDORSEMENT OF RASCH APPLICATIONS

The five interrogation statements were selected to capture the essential elements of the Rasch measurement framework and its implications for the assessment of latent attributes. They address the existence of a distinct Rasch class of measurement, the transformation of subjective responses into measurement, the role of the Rasch logit ratio scale in evaluating therapy impact, the concept of possession of a latent trait and the relationship between Rasch measurement and representational measurement. Taken together, the statements provide a concise profile of the extent to which Rasch concepts are embedded within a target knowledge base.

The results for the three UK institutions are presented first because they occupy a unique position in the history of HTA (Table 1). If Rasch measurement played a role in the development of HTA, evidence of that role would be expected to appear within these knowledge bases.

The findings indicate precisely the opposite. Four of the five statements receive the minimum endorsement probability of 0.05 across all three institutions. These include recognition of the Rasch logit ratio scale, transformation of subjective responses through Rasch rules, the existence of two measurement classes, linear ratio and Rasch logit ratio, and the equivalence between Rasch measurement and representational measurement. Only the statement concerning possession of a latent trait receives endorsement above the floor value, although this remains weak at 0.20 to 0.25.

TABLE 1

ENDORSEMENT OF RASCH MEASUREMENT: YORK CENTRE FOR HEALTH ECONOMICS (CHE), OFFICE OFHEALTH ECONOMICS (OHE) AND NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE (NICE)

CATEGORICAL PROBABILITIES ENDORSEMENT			
STATEMENT	NICE	OHE	CHE
There are only two classes of measurement: linear ratio and Rasch logit ratio	0.05	0.05	0.05
Transforming subjective responses to interval measurement is only possible with Rasch rules	0.05	0.05	0.05
The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits	0.05	0.05	0.05
The outcome of interest for latent traits is the possession of that trait	0.20	0.25	0.25
The Rasch rules for measurement are identical to the axioms of representational measurement	0.05	0.05	0.05

These findings suggest that Rasch measurement was not integrated into the conceptual foundations of HTA during its formative period. While there is limited recognition that latent attributes may exist, there is little evidence that the measurement of those attributes was approached through the Rasch framework. The dominant concern appears to have been valuation of health states rather than measurement of latent constructs.

The Australian results closely mirror the UK findings (Table 2). Across the PBAC and the five Australian HTA research centers, Rasch endorsement probabilities remain at or near floor values for all Rasch statements. Again, the only partial exception is recognition of latent trait possession, where probabilities reach 0.25 for some institutions. Yet even here there is no corresponding recognition of the measurement framework necessary to evaluate possession.

The results presented in Table 1 are among the most revealing findings to emerge from the interrogation of the Australian HTA knowledge base. While previous analyses demonstrated widespread endorsement of utilities, QALYs and reference-case simulation modelling, the Rasch results expose a deeper problem. They suggest not merely a preference for one methodological approach over another, but an almost complete absence of engagement with the problem of measurement itself when the attribute of interest is latent.

The findings are remarkably consistent. Across the PBAC and leading Australian academic research centers in Adelaide, Griffith, Melbourne, Monash and Leeder, endorsement probabilities for the core propositions of Rasch measurement range from 0.05 to 0.10. These are among the lowest endorsement scores recorded in any of the interrogations. The implication is straightforward: the concepts that define Rasch measurement occupy virtually no place within the Australian HTA knowledge base.

TABLE 2

ENDORSEMENT OF RASCH MEASUREMENT: PBAC AND AUSTRALIAN ACADEMIC RESEARCH CENTERS

	ENDORSEMENT CATEGORICAL PROBABILITIES					
STATEMENT	PBAC	ADELAIDE	GRIFFITH	MELBOURNE	MONASH	LEEDER
There are only two classes of measurement: linear ratio and Rasch logit ratio	0.10	0.05	0.05	0.05	0.05	0.05
Transforming subjective responses to interval measurement is only possible with Rasch rules	0.10	0.05	0.05	0.05	0.05	0,05
The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits	0.10	0.05	0.05	0.05	0.05	0,05
The outcome of interest for latent traits is the possession ofthat trait	0.25	0.25	0.05	0.25	0.25	0.05
The Rasch rules for measurement are identical to the axioms of representational measurement	0.10	0.05	0.05	0.05	0.05	0.05

Note: PBAC Pharmaceutica Benefits Advisory Committee; Adelaide: Health Economics Group; Griffith: Centre for Applied Heath Economics; Melbourne: School of Population and Global Health; Monash: Centre for Health Economics; Sydney Leeder Centre for Health Policy

This is a remarkable position for a discipline that places considerable emphasis on patient-reported outcomes, quality of life, treatment satisfaction, symptom burden and other latent attributes. Unlike manifest attributes such as hospital admissions, hospital days or prescription counts, latent attributes cannot be observed directly. Their measurement requires a formal measurement model capable of transforming subjective observations into quantitative measures. Rasch measurement was developed specifically to address this problem. Yet the interrogation results indicate that the Australian HTA community has largely ignored its existence.

Consider the first proposition: there are only two classes of measurement relevant to therapy assessment, linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. Endorsement ranges from 0.05 to 0.10. This finding suggests that the distinction between manifest and latent attributes, fundamental to both measurement theory and scientific inquiry, is largely absent. Instead, latent attributes appear to be treated as though they can be evaluated through ordinal scores, preference values or utility algorithms without establishing whether measurement has occurred.

The second proposition is equally revealing. Transforming subjective responses into interval or ratio measurement is only possible through the application of Rasch rules. Again, endorsement is virtually non-existent. This is important because the dominant approach within HTA assumes that summations of questionnaire responses can be interpreted as though they possess quantitative meaning. Rasch measurement rejects this assumption. Responses to questionnaire items are observations, not measures. Measurement emerges only when those observations satisfy the requirements of a Rasch model. The lack of endorsement suggests that this distinction is not recognized.

The third proposition concerns the assessment of therapy impact for latent traits. Rasch measurement argues that the outcome of interest is change in the possession of the latent attribute, estimated through movement on a Rasch logit scale. Yet endorsement for this proposition again approaches zero. This finding demonstrates that Australian HTA has largely failed to engage with the concept of latent trait possession. Instead, therapy impact is commonly represented through changes in ordinal scores, utility values or composite indices that provide no evidence of lawful measurement.

The fourth proposition performs only marginally better. The statement that the outcome of interest for a latent trait is possession of that trait receives endorsement levels of 0.25 for the PBAC, Adelaide, Melbourne and Monash, but only 0.05 for Griffith and Leeder. Even these higher values remain extremely weak. The implication is that latent attributes are rarely viewed as measurable constructs. Rather, they are treated as collections of responses that can be manipulated arithmetically without first establishing measurement status.

Perhaps the most striking result concerns the final proposition: the Rasch rules for measurement are identical to the axioms of representational measurement. Endorsement again falls between 0.05 and 0.10. This finding is particularly important because it demonstrates that the Australian HTA community has not merely ignored Rasch measurement. It has also failed to recognize the connection between Rasch measurement and the broader tradition of representational measurement theory. The work of Rasch and Wright showed that the Rasch model provides a practical realization of the principles that measurement theorists had long argued were necessary for quantitative science. The absence of endorsement therefore represents a rejection not merely of Rasch, but of the measurement tradition itself.

Taken together, these results paint a troubling picture. Australian HTA has invested enormous intellectual effort in utilities, QALYs, simulation models and cost-effectiveness analysis. Yet when confronted with the only established framework for measuring latent attributes, the response is near-universal indifference. This is particularly ironic because many of the outcomes that HTA claims to value most highly quality of life, symptom burden, functioning, treatment satisfaction and patient benefit are latent attributes. The discipline is therefore deeply interested in latent constructs while simultaneously ignoring the only accepted methodology for measuring them.

The consequences are profound. Without Rasch measurement, latent attributes remain observations rather than measures. Ordinal scores remain ordinal scores. Utility algorithms do not create measurement. Composite indices do not create measurement. Numerical manipulation cannot substitute for measurement. The result is that many patient-reported outcomes used in HTA possess no established measurement status despite being treated as though they support arithmetic operations and comparative value claims.

These findings therefore reinforce the broader conclusion emerging from the Australian interrogations. The issue is not simply that Rasch measurement is absent. The issue is that measurement itself is absent. The near-zero endorsement of Rasch propositions demonstrates a knowledge base that has become preoccupied with arithmetic while neglecting the conditions necessary for measurement. In a discipline concerned with therapy impact, this omission is extraordinary. It leaves Australian HTA in the position of discussing latent outcomes without the means to measure them. The result is a framework characterized not by quantitative science, but by the continuing substitution of numerical manipulation for measurement.

IMPLICATIONS FOR HTA TEACHING AND RESEARCH: ARITHMETIC CHAOS

The invisibility of Rasch measurement has consequences that extend far beyond a technical dispute over questionnaire design or patient-reported outcomes. It points to a much deeper problem within HTA teaching and research: the acceptance of arithmetic operations without prior demonstration that the entities entering those operations possess lawful measurement properties. The result is what may be described as arithmetic chaos.

Arithmetic chaos emerges when numerical entities are combined, transformed and manipulated without regard to the constraints imposed by scales of measurement and the axioms of representational measurement. The issue is not whether calculations can be performed. Modern computing systems can perform almost limitless calculations. The issue is whether the resulting arithmetic operations possess recognized quantitative meaning. If the underlying entities are not lawful measures, the outputs derived from them cannot acquire validity merely through increasingly sophisticated analytical techniques.

The significance of the Rasch findings is that they reveal an intellectual environment where the measurement of latent attributes is largely absent from the curriculum itself. Across HTA teaching and research programs, students are introduced to utilities, QALYs, Markov simulation models, probabilistic sensitivity analysis, discounting procedures and cost-effectiveness ratios. Yet there is little evidence that they are introduced to the problem that originally motivated Rasch: how subjective observations can be transformed into measures. The consequence is that ordinal and composite ordinal structures are routinely treated as though they possess the properties necessary for multiplication, aggregation and ratio construction.

This omission is important because the entire reference-case framework depends upon arithmetic operations applied to subjective constructs. Health-state descriptions generate preference scores; preference scores are converted into utilities; utilities are multiplied by time to create QALYs; QALYs enter simulation models together with costs and assumptions; simulation models generate cost-effectiveness ratios. At each stage arithmetic operations proceed as though lawful measurement has already been established. Yet the interrogation results suggest that the measurement framework necessary to justify these operations is largely invisible within HTA itself.

The result is institutionalized arithmetic chaos. Students inherit a framework in which arithmetic is accepted as legitimate by convention rather than justified through measurement theory. Researchers refine assumptions, expand models and develop increasingly sophisticated simulations, while the prior question remains unasked: do the numerical entities entering these models possess lawful measurement properties? Without an answer to that question, HTA risks becoming a discipline in which numerical complexity substitutes for measurement.

The irony is striking. For more than sixty years Rasch measurement has provided a coherent solution to the problem of transforming subjective observations into measures. Yet within HTA teaching and research, that solution remains largely invisible. The consequence is a curriculum that reproduces arithmetic chaos across successive generations of researchers and practitioners while remaining largely unaware that an alternative measurement framework exists.

ONLY TWO MEASURES

The solution to the invisibility of Rasch measurement and the institutionalization of arithmetic chaos is remarkably simple. It does not require refinement of the reference-case framework, more sophisticated utility algorithms, larger simulation models or increasingly elaborate sensitivity analyses. The solution is to abandon the reference-case architecture entirely and return HTA to the standards of measurement and normal science.

The starting point is recognition that there are only two lawful forms of measurement relevant to therapy impact assessment. The first is the linear ratio measure required for manifest attributes. These are directly observable phenomena such as hospital admissions, hospital days, physician visits, persistence, therapy switching, emergency department utilization and other objectively recorded events. Such attributes possess natural quantitative structures and can support lawful arithmetic operations provided the requirements of ratio measurement are satisfied.

The second is the Rasch logit ratio measure required for latent attributes. These include symptom burden, quality of life, treatment satisfaction, need fulfilment and other constructs that cannot be observed directly but are inferred from patient responses. For these attributes, questionnaire responses are not measures. They are observations from which a measure may be constructed if, and only if, the Rasch requirements for measurement are satisfied. The outcome of interest is not an aggregate score but possession of the latent attribute located on an invariant Rasch ruler.

This distinction immediately transforms the purpose of HTA. Instead of constructing utilities, QALYs and cost-effectiveness ratios through chains of inadmissible arithmetic operations, attention shifts to explicit attribute claims. The question becomes straightforward: what attribute is expected to change as a consequence of therapy intervention, and how should that change be measured?

Each attribute claim then requires a supporting protocol. The protocol identifies the target population, specifies the attribute of interest, describes the measurement instrument, defines the observation period and states the expected magnitude and direction of change. Most importantly, the protocol establishes the conditions under which the claim may be challenged.

Falsification therefore replaces simulation. Rather than generating hypothetical outcomes decades into an uncertain future, HTA returns to the standards of empirical science. Claims are exposed to observation, replication and possible failure. Evidence emerges through measurement rather than through assumption-driven modelling. The transition is therefore not merely methodological. It represents a return to a framework where quantitative claims derive their legitimacy from lawful measurement, explicit protocols and empirical challenge rather than from the institutionalized arithmetic chaos of the reference-case model.

RASCH AND THE RECONSTRUCTION OF HTA

The significance of Rasch measurement extends well beyond the technical challenge of transforming subjective observations into measures. Rasch represents an alternative vision of what HTA might have become had measurement rather than valuation of health states occupied the center of its intellectual development. Indeed, it is difficult to avoid the conclusion that the invisibility of Rasch within HTA deprived the discipline of precisely the framework it required to avoid measurement inversion and the arithmetic chaos that subsequently emerged.

The contrast between Rasch measurement and the reference-case framework is striking. Rasch begins with an attribute. It asks whether the attribute is latent, whether it can be measured and whether observations satisfy the requirements necessary to support measurement. Instrument development is driven by explicit rules governing item selection, unidimensionality, invariance and model fit. Measurement is never assumed. It must be demonstrated. Arithmetic follows only after the measurement properties of the instrument have been established.

The reference-case framework evolved in the opposite direction. Rather than beginning with measurement, it began with valuation. Health-state descriptions were assigned preference values, preference values became utilities and utilities were subsequently incorporated into QALYs, simulation models and cost-effectiveness ratios. Arithmetic operations proliferated while the question of measurement largely disappeared. The result was a framework that generated increasingly elaborate numerical outputs while remaining detached from the standards governing lawful quantitative inference.

What makes this omission particularly important is that Rasch embodies many of the principles associated with normal science. Claims are explicit. Measurement properties are transparent. Model fit is continually assessed. Instruments may fail and require revision. Evidence accumulates through replication and empirical challenge rather than through increasingly complex assumptions. In this respect Rasch is not simply a measurement model. It is an example of a scientific framework where measurement, evidence and falsification remain central.

The implications for HTA are profound. A reconstructed HTA would not focus upon utilities, QALYs or lifetime simulation models. Instead, it would begin with attribute claims. Manifest attributes would require linear ratio measures. Latent attributes would require Rasch logit ratio measures. Claims would be linked to protocols, exposed to empirical evaluation and subject to falsification. Therapy assessment would therefore return to measurable outcomes rather than simulated estimates derived from chains of inadmissible arithmetic operations.

The invisibility of Rasch within HTA is therefore more than a historical curiosity. It represents a missed opportunity in the development of the discipline. For more than sixty years a coherent framework existed for transforming subjective observations into lawful measures, yet HTA largely ignored it. The consequence was the emergence of a valuation-based architecture characterized by measurement inversion and arithmetic chaos. Rasch therefore stands not merely as an alternative methodology but as the exemplar of a reconstructed HTA founded upon measurement, explicit claims, empirical challenge and the continuing evolution of objective knowledge.

CLOSURE AND TRANSITION

Recognition of closure immediately raises a practical question: what standards are required if Australian HTA is to transition from a primitive framework defined by measurement inversion and arithmetic chaos toward one consistent with quantitative science? Criticism alone is insufficient. Once arithmetic detached itself from lawful measurement, closure became inevitable. The issue now is reconstruction. The following standards do not represent refinements to the existing reference-case architecture. They define the conditions necessary for replacement.

1. Measurement must precede arithmetic

This principle is the point from which all reconstruction begins. For decades HTA largely assumed that attaching numbers to observations automatically created quantitative evidence. Yet arithmetic operations cannot create measurement properties where they do not already exist. Multiplication, averaging and aggregation acquire meaning only after lawful measurement structures have been demonstrated. The failure to recognize this sequence created the conditions for measurement inversion itself.

2. Scale classification becomes a prerequisite for evidence generation

Not all numerical scales possess the same properties. Nominal, ordinal, interval and ratio scales support different forms of interpretation and different arithmetic operations. Historically these distinctions often disappeared within HTA practice as scores and utilities were treated as quantities without first asking what type of scale existed. Reconstruction requires explicit recognition of scale structure before claims can be advanced.

3. Ratio measures occupy a unique role in therapy evaluation
Only ratio measures support unrestricted arithmetic because they preserve meaningful quantitative relationships and possess identifiable origins. Under conventional HTA, ratio assumptions were frequently imposed upon structures lacking these properties. Future therapy claims therefore require lawful ratio measures rather than numerical approximations presented as quantities.

4. All therapy impact claims must concern attributes
Therapies do not act upon utilities, scores or simulation outputs. Therapies influence attributes within patients and populations. Hospital utilization, persistence, symptom burden and treatment satisfaction represent examples. Reconstruction therefore begins by asking a simple question largely absent from conventional HTA: what attribute is expected to change?

5. Manifest and latent attributes must be distinguished
One of the major consequences of measurement inversion was failure to distinguish between directly observed and inferred phenomena. Manifest attributes such as physician visits or hospitalization differ fundamentally from latent constructs such as symptom burden or need fulfilment.

6. Manifest attributes require linear ratio measures
Observable phenomena often possess natural quantitative structures. Hospital days, persistence behavior and switching rates provide examples where arithmetic may be meaningful. Manifest claims therefore require explicit ratio measurement structures. Quantitative interpretation follows from the properties of the attribute itself rather than numerical conventions imposed upon observations.

7. Latent attributes require Rasch logit ratio measures
Latent constructs create a different challenge. Responses to questionnaires are not measures simply because they have numbers attached to them. Scores are observations, not quantities. Reconstruction requires recognition that latent attributes demand Rasch procedures capable of creating lawful logit ratio structures and evaluating possession. The outcome of interest becomes possession of the latent attribute rather than change in arbitrary scores.

8. Attribute claims require explicit justification
Under conventional HTA, claims frequently emerged as outputs from analytical systems rather than propositions requiring prior explanation. Utilities generated QALYs and models generated cost-effectiveness estimates, yet the rationale for selecting particular outcomes often remained unclear. Reconstruction requires a different approach. Each attribute claim should provide an explicit justification explaining why therapy is expected to influence the attribute, what evidence supports this expectation and why the anticipated change possesses clinical relevance. Claims should therefore emerge from a coherent explanatory framework rather than analytical convenience or historical precedent.

9. Attribute claims require protocols
Protocols become the mechanism through which claims acquire scientific standing. Under reference-case systems assumptions frequently disappeared within simulation structures and remained insulated from direct challenge. Reconstruction requires every attribute claim to be linked prospectively to a protocol specifying the target population, observation period, expected treatment effects and evaluation procedures. Protocols therefore become scientific instruments rather than administrative documents. They establish the framework through which claims move from expectation to evidence.

10. Attribute claims must be empirically evaluable
Claims cannot remain hypothetical projections extending decades into uncertain futures. A claim acquires scientific meaning only when evidence capable of supporting or challenging it can be generated. Reconstruction therefore requires attribute claims to specify measurable endpoints, defined populations and observation schedules capable of prospective assessment. Evaluation becomes a continuing process rather than a one-time analytical exercise completed at product launch.

11. Falsification becomes a requirement
Scientific progress requires more than confirmation; it requires vulnerability to challenge. Attribute claims should therefore specify the conditions under which they may fail. If anticipated effects are absent or treatment outcomes diverge substantially from expectations, claims should be modified or rejected. Exposure to challenge replaces protection of assumptions. Claims incapable of failure cease to function as scientific propositions.

12. HTA returns to the standards of normal science
Taken together these standards do not create a new ideology. They restore principles long familiar within quantitative science: measurable attributes, lawful measurement structures, explicit claims and continuing empirical challenge. Numerical storytelling gives way to measurable propositions. The objective is straightforward: reconnect HTA with the continuing evolution of objective knowledge. HTA therefore re-enters a framework where evidence grows through replication, criticism and empirical learning rather than through increasingly elaborate assumptions.

CONCLUSION

The purpose of this paper has not been to criticize Rasch's absence from HTA for its own sake. Rather, it has been to understand how a discipline so dependent upon patient-reported outcomes, quality-of-life claims and other latent attributes could develop while largely ignoring the one measurement framework specifically designed to transform subjective observations into lawful measures. The interrogation results suggest that the invisibility of Rasch was inherited from the intellectual foundations of HTA itself and subsequently reproduced through teaching, research and policy development in Australia. The absence was never questioned.

Yet there is an encouraging aspect to this conclusion. The invisibility of Rasch is not a permanent condition. Unlike many scientific controversies, the solution already exists. For more than sixty years Rasch measurement has provided a coherent and operational framework for transforming subjective responses into invariant measures, evaluating latent attribute possession and supporting lawful quantitative inference. The challenge is therefore not to invent a new methodology but to recognize and apply one that has long been available.

The implications for HTA are potentially transformative. Bringing Rasch in from the cold would do more than improve patient-reported outcome assessment. It would redirect HTA toward measurable attributes, explicit claims, protocol-driven evaluation and empirical challenge. It would reconnect therapy assessment with the standards of representational measurement and restore the principle that measurement must precede arithmetic. Most importantly, it would replace the closed architecture of valuation and simulation with a framework capable of supporting the continuing evolution of objective knowledge.

The significance of these findings lies not in any individual statement but in the remarkable consistency of the endorsement structure. Across institutions separated by geography, policy responsibilities and academic focus, the same pattern appears repeatedly. Rasch measurement is not merely underutilized. It appears largely invisible. The Australian results therefore support the hypothesis that this invisibility was inherited from the UK institutions that shaped the development of modern HTA. The issue is not that Rasch was considered and rejected. Rather, the evidence suggests that Rasch measurement remained outside the conceptual boundaries of the Australian HTA framework from its inception and has remained there ever since. The authorities were not given a chance to consider its role.

This is a rich research agenda. Questions of latent attribute development, therapy impact assessment, instrument construction, claim evaluation, replication and falsification offer opportunities for decades of productive inquiry. Rather than defending increasingly elaborate simulation models, researchers would be able to focus on the development and evaluation of lawful measures and empirically testable claims. HTA would become not a closed system of numerical storytelling but an open scientific enterprise committed to measurement, evidence and learning.

The future of HTA therefore need not be defined by measurement inversion and arithmetic chaos. Rasch measurement provides a pathway toward reconstruction. The opportunity now is to move it from the margins of HTA discourse to the center of a new research and teaching environment consistent with the standards of normal science and the continuing evolution of objective knowledge.

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

[i] Stevens, S. S. 1946. “On the Theory of Scales of Measurement.” *Science* 103 (2684): 677–680.

[ii] Rasch, Georg. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research (Danmarks Paedagogiske Institut), 1960. [Reprinted in expanded form by the University of Chicago Press in 1980 with a foreword and afterword by Benjamin D. Wright].

[iii] Wright B. Misunderstanding the Rasch Model. *J Educational Measurement* 1977;14 (1): 75–96.

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

[v] United Kingdom: The Office of Health Economics – The inversion of representational measurement, Logit Working Paper No 25. https://maimonresearch.com/logit-working-paper-no-25/

[vi] United Kingdom: the York Centre for Health Economics – A canonical knowledge base interrogation and the absence of measurement, Logit Working Paper No 780. https://maimonresearch.com/logit-working-paper-no-780-april-2026/

[vii] United Kingdom: The National Institute for Health and Care Excellence (NICE) and the absence of representational measurement. Logit Working Paper No 701. https://maimonresearch.com/logit-working-paper-no-701-march-2026-rev-8/

[viii] Australia: Pharmaceutical Benefits Advisory Committee (PBAC) – decisions without measurement. Logit Working Paper No 34., https://maimonresearch.com/logit-working-paper-no-34/

[ix] Australia: the absence of representational measurement and the Melbourne School of Population and Global Health (MSPGH) University of Melbourne. Logit Working Paper No 724. https://maimonresearch.com/logit-working-paper-no-724-february-2026/

[x] Australia: The absence of representational measurement and the Centre for Health Economics (CHE) Monash University, https://maimonresearch.com/logit-working-paper-no-722-february-2026/

[xi] Australia: The absence of representational measurement and the Health Economics Goup (HEG) University of Adelaide. https://maimonresearch.com/logit-working-paper-no-723-february-2026/

[xii] Australia: The absence of representational measurement and the Centre for Applied Health Economics (CAHE) Griffith University , https://maimonresearch.com/logit-working-paper-no-727-february-2026-2/

[xiii] Australia: The absence of representational measurement and the Leeder Centre for Health Policy, University of Sydney Log Working Paper No 791. https://maimonresearch.com/logit-working-paper-no-791-april-2026/