

**RESEARCH LLC**

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE  
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN  
HEALTH TECHNOLOGY ASSESSMENT**

**GERMANY: CURRICULUM INVERSION IN HEALTH  
TECHNOLOGY ASSESSMENT**

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

**LOGIT WORKING PAPER No 1470 JULY 2026**

**[www.maimonresearch.com](http://www.maimonresearch.com)**

**Tucson AZ**

## ABSTRACT

*This paper examines whether the German health technology assessment (HTA) educational environment provides the scientific foundations required for lawful quantitative claims regarding therapy impact. It complements an earlier assessment of measurement inversion by examining curriculum inversion: the failure to teach the principles of measurement before introducing students to the analytical methods that depend upon them. The assessment interrogates the German HTA curriculum knowledge base, defined as the publicly accessible educational, methodological, research, and policy materials that collectively shape HTA teaching and professional development. A large language model interrogation was undertaken using ten canonical curriculum statements representing the minimum scientific competencies expected of graduates entering HTA practice.*

*The results indicate a consistent pattern of curriculum inversion. Strong endorsement is found only for the recognition of manifest empirical observation, while the concepts that underpin measurement-based science receive little support. These include target attribute specification, the principal scales of measurement, representational measurement, unidimensionality, the distinction between manifest and latent attributes, ratio measurement, Rasch measurement for latent constructs, and the requirement that therapy-impact claims be prospectively evaluable and falsifiable. The findings suggest that students are introduced to economic evaluation, comparative effectiveness, utility assessment, and modelling without first acquiring the conceptual framework necessary to determine whether these methods support scientifically defensible quantitative claims.*

*The implications extend beyond university teaching. Curriculum inversion provides the educational mechanism through which measurement inversion is reproduced across successive generations of researchers, reviewers, HTA agencies, journals, and policy makers. Germany therefore faces not merely an educational issue but a scientific one. Professional competence in HTA should include mastery of the principles of representational measurement before competence in economic evaluation or simulation modelling can be claimed. The paper concludes that Germany should undertake a national assessment of curriculum inversion to determine whether its educational foundations satisfy the standards required for quantitative science and whether reconstruction of HTA education is necessary to support evaluable, replicable, and falsifiable claims regarding therapy impact.*

## INTRODUCTION

The starting point for any reconstruction of health technology assessment (HTA) in the UK must be an understanding of the scale and consistency of the false measurement inversion that characterizes the field. Over the past several months, a series of large language model (LLM) interrogations has been undertaken to evaluate HTA knowledge bases across multiple jurisdictions. These interrogations have encompassed national reimbursement agencies, academic research centers, professional organizations, journals, pharmacy schools and HTA teaching programs in Australia, New Zealand, the United Kingdom, Europe, Singapore and the United States. Although the institutions differ in structure, mission and geographical location, the results have been remarkably consistent.

The concept of measurement inversion is straightforward. In the physical and social sciences, measurement precedes arithmetic. Before quantities can be added, multiplied, transformed or modelled, the measurement properties of those quantities must first be established. HTA in Canada reverses this sequence. Numerical operations are routinely undertaken without demonstrating that the quantities involved satisfy the requirements of measurement. Arithmetic is treated as a substitute for measurement rather than a consequence of measurement. The result is a framework in which numerical manipulation takes precedence over the validation of the quantities being manipulated.

The large language model (LLM) interrogations reveal this inversion repeatedly <sup>i</sup>. Across institutions there is little awareness of scales of measurement, the distinction between ordinal, interval and ratio scales, the role of unidimensionality, the axioms of representational measurement, dimensional homogeneity, or the distinction between manifest and latent attributes. At the same time, there is widespread endorsement of analytical frameworks that depend upon precisely those concepts. Utilities are treated as though they possess interval or ratio properties. Quality-adjusted life years are accepted as quantitative measures despite the absence of demonstrable ratio-scale foundations. Reference-case simulation models are employed to generate lifetime cost-effectiveness claims without establishing the measurement properties of the underlying inputs. The pattern is universal.

The significance of these findings lies not merely in the demonstration of measurement failure but in the realization that measurement inversion is now institutionalized. It is embedded in teaching programs, methodological guidance, research publications and policy frameworks. Canada is no exception. Researchers entering the field encounter an intellectual environment in which the requirements of measurement are rarely discussed and almost never presented as prerequisites for quantitative claims. Consequently, the acceptance of measurement inversion becomes self-reinforcing. Successive generations of researchers inherit analytical frameworks without being introduced to the measurement principles required to evaluate them. It is possible for a professional career to pass without knowledge of the scales of measurement or the axioms of representational measurement <sup>ii iii</sup>.

It is worth noting the complete absence of Rasch measurement from the curriculum profiles of the Canada HTA research centers <sup>iv</sup>. This omission is particularly significant because Rasch measurement represents the only established framework that provides the necessary and sufficient conditions for transforming observations relating to a latent attribute into a measure. The central issue is not statistical sophistication but measurement itself. Latent attributes such as symptom burden, functional status, treatment satisfaction, need fulfilment and quality of life cannot be observed directly. They require a measurement model capable of estimating possession of the attribute while demonstrating unidimensionality, invariance and appropriate scale properties. Rasch measurement was developed specifically to meet these requirements. Its absence suggests that faculty and students are not exposed to the concept of latent attribute possession, nor to the scientific challenge of constructing measures from observations. Instead, there is a tendency to treat questionnaire scores, indexes and composite summaries as though they were measures in their own right.

The result is that one of the most important developments in measurement theory over the past 80 years is effectively invisible within Canada HTA education. Without exposure to Rasch measurement, faculty and students are unlikely to appreciate that latent attributes require a fundamentally different approach to measurement from manifest attributes, or that quantitative claims regarding latent phenomena require a demonstrable measurement framework before arithmetic operations can be justified.

This observation raises an obvious question. If measurement inversion is so widespread and persistent, where does it originate? The answer cannot be found solely in research practice or policy guidance. Attention must also be directed toward education. If the concepts necessary to recognize measurement failure are absent from the curriculum, then measurement inversion becomes the expected outcome rather than an isolated error. Researchers cannot be expected to evaluate measurement claims and the unique contribution of ratio measurement if they have never been introduced to scales of measurement, representational measurement, ratio scales, unidimensionality, or the distinction between manifest, latent attributes and even Rasch measurement <sup>v</sup>

## **STANDARDS FOR MEASUREMENT**

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

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

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

These requirements are formalized in the axioms of representational measurement. Representational measurement provides the scientific framework that links empirical observations to numerical representations. Its purpose is to ensure that numerical assignments preserve the structure of the attribute being measured. Only when this correspondence is demonstrated can

arithmetic operations be regarded as meaningful. The axioms of representational measurement therefore establish the conditions under which quantitative claims can be considered scientifically legitimate.

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

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

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

## **CURRICULUM EXPECTATIONS**

A curriculum intended to prepare students for health technology assessment should provide more than instruction in economic evaluation, evidence synthesis, utility assessment, and simulation modelling. It should equip students with the conceptual tools necessary to evaluate the scientific legitimacy of quantitative claims.

At a minimum, students should be introduced to the role of attributes in therapy assessment and the requirement that the target attribute be specified before evaluation begins. They should understand the principal scales of measurement and the implications these have for admissible arithmetic. They should be familiar with the distinction between nominal, ordinal, interval, and ratio scales and appreciate that different scales support different forms of analysis.

A scientifically grounded curriculum should also introduce students to representational measurement and its role in determining whether numerical assignments constitute valid measures. This includes an understanding of the relationship between measurement and arithmetic, the importance of unidimensionality, and the conditions required for quantitative claims.

Students should further be expected to recognize the distinction between manifest and latent attributes. Observable outcomes such as survival, hospital admissions, adverse events, and treatment persistence present different measurement challenges from latent constructs such as pain, fatigue, quality of life, functioning, and patient satisfaction. A curriculum should therefore provide an understanding of the different measurement frameworks required for these two classes of attributes.

Finally, students should be exposed to the principles of empirical evaluation and falsification. Quantitative claims should not be treated as self-validating simply because they are numerical. Rather, students should understand that scientific claims must be capable of evaluation, replication, and potential refutation through observation and evidence.

These expectations do not represent advanced or specialist topics. They constitute the foundational knowledge required for any discipline that seeks to support quantitative claims regarding therapy impact. Their presence would indicate a curriculum grounded in measurement and scientific inquiry. Their absence would suggest that students are being trained to apply analytical techniques without first acquiring the conceptual framework necessary to evaluate whether the resulting claims are scientifically defensible.

## **CURRICULUM INVERSION**

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

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

The importance of curriculum inversion extends beyond education. Universities and research centers are responsible for training the individuals who subsequently populate HTA agencies, reimbursement committees, consulting organizations, professional societies, and academic departments. If the concepts necessary to evaluate quantitative claims are absent from the curriculum, they will also be absent from research, policy, and practice. The omissions identified in educational programs therefore become embedded throughout the wider HTA environment. What begins as a curriculum deficiency ultimately becomes a disciplinary norm.

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

This is why curriculum inversion provides an important explanation for the persistence of the reference-case paradigm. Utilities, QALYs, cost-effectiveness ratios, and simulation models have survived not because their measurement foundations were demonstrated, but because the concepts required to challenge those foundations were largely absent from the educational environment. Curriculum inversion reproduces measurement inversion. Together they explain how a framework built on arithmetic before measurement could become established, institutionalized, and defended for more than four decades.

## **DEFINING THE KNOWLEDGE BASE**

The first step in any LLM interrogation is to define the knowledge base to be interrogated. The validity of the interrogation depends upon ensuring that the knowledge base reflects the information environment that shapes teaching, research, professional training, and policy practice. A knowledge base is more than a collection of documents. It represents the accumulated concepts, assumptions, methods, definitions, and analytical frameworks communicated through national educational, methodological, research, and policy activities. It is the intellectual environment within which students, researchers, agency staff, consultants, and professional practitioners acquire their understanding of HTA.

In the present assessment, the German HTA research and training knowledge base is defined as the publicly accessible body of materials that contributes to HTA education, research, methodological development, and professional practice in Germany. This includes university program descriptions, course outlines, doctoral training resources, research-center materials, methodological guidance, IQWiG and G-BA documents, EUnetHTA-related materials, seminar and workshop content, faculty publications, professional reports, conference presentations, policy briefs, journal articles, and other resources through which HTA concepts are communicated and reinforced. Taken together, these materials provide a representation of the concepts and analytical frameworks that students, researchers, and professional staff are most likely to encounter within the German HTA environment.

The importance of defining a national research and training knowledge base extends beyond formal course content. National HTA systems do more than apply methods. They define what counts as legitimate evidence, what questions are regarded as relevant, what assumptions are accepted, what analytical techniques are considered standard, and what competencies are expected of future practitioners. Over time, these priorities shape the intellectual culture of the discipline. Concepts that are consistently emphasized become embedded in professional practice. Concepts that are neglected disappear from education, research, agency guidance, journal expectations, and policy assessment.

A German HTA curriculum designed to support scientific evaluation of therapeutic interventions would be expected to expose students and practitioners to the foundational principles required for quantitative reasoning. These include specification of the target attribute, the principal scales of measurement, representational measurement, admissible arithmetic, unidimensionality, dimensional homogeneity, the distinction between manifest and latent attributes, ratio measurement, Rasch measurement for latent attributes, and the requirement that therapy-impact claims be capable of empirical evaluation, replication, and potential falsification. These concepts are not specialist additions to HTA. They are the intellectual foundations upon which all quantitative claims ultimately depend.

The objective of the present interrogation is therefore not to evaluate a single German university, research center, agency, course, faculty member, or teaching style. Rather, it is to assess the broader national intellectual environment represented by the German HTA research and training knowledge base. The question is straightforward: to what extent are the concepts required for measurement-based therapy assessment present, absent, or weakly represented within the German educational, research, methodological, and policy environment? The answer provides an indication of whether future German HTA researchers, analysts, agency staff, and decision makers are likely to encounter the conceptual foundations necessary to evaluate quantitative claims, or whether they are being trained primarily in the application of established analytical techniques. In this respect, the German HTA research and training knowledge base provides a window into the scientific foundations of HTA education and professional formation in Germany.

## **INTERROGATING THE CURRICULUM KNOWLEDGE BASE**

The purpose of the present interrogation is not to determine what a large language model "believes." Large language models possess no beliefs, opinions, or independent understanding. Their value lies in their ability to interrogate and summarize the conceptual content of a defined knowledge base. The objective is therefore to examine the intellectual environment represented within a curriculum and to determine which concepts are present, absent, weakly represented, or strongly reinforced.

In the context of curriculum assessment, the knowledge base is defined as the totality of publicly accessible materials through which a research center communicates educational and methodological content. These materials include curriculum descriptions, course outlines, learning objectives, training programs, methodological guidance documents, workshop and seminar content, faculty publications, doctoral training resources, technical reports, conference presentations, and other educational materials associated with the institution. Taken together, these

sources provide a representation of the concepts and analytical frameworks that students, researchers, and professional staff are most likely to encounter.

The interrogation therefore focuses not on individual instructors or specific courses but on the broader intellectual environment that the curriculum creates and sustains. A curriculum functions as a mechanism for transmitting knowledge across successive generations of students and researchers. Concepts that are consistently emphasized become embedded within professional practice. Concepts that are absent or weakly represented are unlikely to become part of the analytical framework employed by future practitioners. The curriculum knowledge base therefore provides an important indicator of the intellectual foundations of a discipline.

For the purposes of the present assessment, the objective is straightforward: to determine the extent to which the curriculum knowledge base recognizes and reinforces the concepts required for a measurement-based approach to health technology assessment. Particular attention is given to attributes, measurement scales, representational measurement, admissible arithmetic, unidimensionality, the distinction between manifest and latent attributes, ratio measurement, Rasch measurement, and the requirement that scientific claims be evaluable and potentially falsifiable. These concepts constitute the foundational framework necessary for evaluating the validity of quantitative claims.

The interrogation proceeds through a series of canonical statements representing concepts that would be expected to appear within a curriculum grounded in measurement science. The resulting endorsement probabilities provide an indication of the extent to which these concepts are embedded within the curriculum knowledge base. High endorsement probabilities suggest that the concept is visible and reinforced within the educational environment. Low endorsement probabilities suggest that the concept is absent, weakly represented, or largely ignored.

The significance of this approach is that it allows the curriculum to be evaluated as a coherent intellectual system rather than as a collection of individual courses. The question is not whether students are taught contemporary HTA methods. The question is whether they are taught the concepts necessary to evaluate the scientific legitimacy of those methods. The resulting profile provides a measure of the extent to which the curriculum supports, neglects, or potentially undermines the foundations of measurement-based scientific inquiry.

## **TOOLS FOR INTERROGATION**

The identification of measurement inversion across HTA research centers, reimbursement agencies and academic programs raises an obvious question: where does this inversion originate?

The objective of curriculum interrogation differs from that of previous HTA knowledge-based practice assessments. Earlier interrogations focused on whether institutions recognized the requirements of representational measurement and the standards necessary for quantitative claims. Curriculum interrogation asks a different question. Are faculty, students and researchers exposed to the concepts necessary to understand and apply those standards? The focus shifts from methodological outputs to educational inputs. Rather than examining what faculty, students and researchers do, attention is directed to what they are taught and what they know.

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

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

These statements are:

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

These ten statements form a logical sequence:

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

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

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

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

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

## **ENDORSEMENT OF CURRICULUM INVERSION**

The endorsement profile is clear (Table 1). The German HTA research and training knowledge base gives strong recognition to manifest empirical evidence but weak recognition to the scientific foundations of measurement. This is consistent with Germany's institutional orientation toward evidence of benefit, comparative clinical assessment, patient-relevant outcomes, and statutory decision-making. The German system is highly developed as a framework for evidence appraisal and reimbursement decision-making. It is not, however, visibly structured as a curriculum in representational measurement.

**TABLE 1: CURRICULUM CONTENT ENDORSEMENT - GERMANY**

CANONICAL STATEMENT	CATEGORICAL PROBABILITY	NORMALIZED LOGIT
An attribute is the specific outcome of interest in a therapy assessment	0.20	-1.40
Every therapy assessment begins with specification of the target attribute	0.15	-1.70
The principal scales of measurement (nominal, ordinal, interval and ratio)	0.10	-2.20

have different properties and support different forms of analysis		
The measurement status of a target attribute must be established before quantitative claims can be advanced	0.05	-2.50
The axioms of representational measurement underpin quantitative claims	0.05	-2.50
Attributes must be demonstrated to be unidimensional before measurement is possible	0.10	-2.20
A manifest attribute is directly observable and capable of supporting empirical observation	0.85	+1.70
A latent attribute is not directly observable and requires a measurement model to estimate possession of the attribute	0.20	-1.40
Manifest and latent attributes require different forms of ratio measurement	0.05	-2,50
Therapy impact claims must be falsifiable	0.15	-1.70

The strongest endorsement is for the statement that a manifest attribute is directly observable and capable of supporting empirical observation. This is unsurprising. German HTA practice places substantial emphasis on clinical evidence, comparative benefit, patient-relevant outcomes, morbidity, mortality, adverse events, and quality of life. IQWiG's methodological materials describe assessment procedures as frameworks for evaluating medical interventions, with the specific assessment depending on the research question and the available scientific evidence. G-BA's early benefit assessment process similarly centers on the evaluation of additional benefit of new medicines. In this sense, the German knowledge base is strongly empirical. It recognizes that observable outcomes matter.

Yet empirical orientation is not the same as measurement science. A curriculum may train students to value randomized trials, comparative effectiveness, registry evidence, subgroup analysis, and patient-relevant outcomes while still failing to teach the measurement principles required for quantitative claims. This distinction is central. The fact that an outcome is observed does not by itself establish the scale properties of the resulting variable. Counts, durations, ratings, classifications, ordinal scores, and composite indices do not support the same forms of arithmetic. Students must therefore be taught not merely to identify outcomes, but to determine their measurement status.

The first two statements concern attributes. The endorsement probabilities of 0.20 and 0.15 indicate limited recognition that therapy assessment begins by specifying the attribute of interest. German HTA guidance and training environments undoubtedly refer to outcomes, endpoints, patient relevance, clinical benefit, and evidence requirements. However, the curriculum-facing knowledge base does not appear to begin from the more fundamental measurement question: what

is the target attribute whose change is being assessed? Survival time, symptom severity, disease progression, treatment persistence, functional ability, pain, fatigue, cognition, caregiver burden, and quality of life are not interchangeable. Each refers to a distinct attribute. Each requires its own measurement strategy. If this specification is not the first step, then subsequent analysis risks becoming method-driven rather than attribute-driven.

The third statement concerns the principal scales of measurement: nominal, ordinal, interval, and ratio. The endorsement probability of 0.10 indicates weak recognition. This is a critical finding because scale type determines permissible analysis. Nominal classifications support counting and classification. Ordinal scores support ranking but not meaningful differences, multiplication, or division. Interval measures support differences but lack a true zero. Ratio measures alone support proportional comparisons and the full range of arithmetic operations. If students are not taught these distinctions before learning economic evaluation, benefit assessment, or modelling, then they are not being given the tools required to determine whether the arithmetic in HTA is scientifically admissible.

The fourth statement receives an endorsement probability of 0.05: the measurement status of a target attribute must be established before quantitative claims can be advanced. This is the central curriculum inversion finding. Contemporary HTA education commonly introduces students to evidence synthesis, comparative benefit, economic evaluation, real-world evidence, utility values, quality-of-life assessment, and decision analysis before asking whether the quantities being analyzed are measures. The German case appears no different. The research and training knowledge base emphasizes assessment, evidence, decision-making, and policy relevance; it gives little visible priority to the prior question of measurement status.

The fifth statement, that the axioms of representational measurement underpin quantitative claims, also receives 0.05. This is not a peripheral omission. Representational measurement provides the formal scientific basis for assigning numbers to empirical attributes in ways that preserve relational structure. Without representational measurement, numerical values may still be generated, but their status as measures remains unproven. A curriculum that omits representational measurement deprives students of the framework needed to determine whether HTA numbers are scientific measures or administrative constructions. This omission is particularly important when students encounter utilities, preference scores, composite outcomes, quality-of-life instruments, or modelled estimates.

The sixth statement concerns unidimensionality. The endorsement probability of 0.10 indicates weak recognition that an attribute must be demonstrated to be unidimensional before measurement is possible. This issue is central to HTA because many widely used instruments combine multiple dimensions into single scores. Health-related quality of life, functional status, symptom burden, patient-reported outcomes, caregiver impact, and disease burden are often treated as if they were measurable quantities. Yet if multiple attributes are combined into a composite score, the result may summarize information but does not automatically constitute measurement. Without unidimensionality, no single attribute has been measured. If this is not taught, students may become fluent in scoring instruments while lacking the scientific basis to determine whether those scores are measures.

The seventh statement, as noted, is the exception. The German knowledge base strongly recognizes manifest empirical observation. This reflects the clinical and evidentiary orientation of German HTA. It also points to an opportunity. Germany may be better positioned than some reference-case systems to reconstruct HTA around manifest therapy-impact claims because its benefit-assessment tradition already gives prominence to observable clinical outcomes. Manifest attributes such as survival time, hospital days, adverse event counts, discontinuations, exacerbations, relapses, and resource-use units can support lawful quantitative claims when appropriately defined on ratio scales. However, this potential is not realized unless the curriculum explicitly teaches manifest attributes as a measurement category governed by scale requirements and admissible arithmetic.

The eighth statement concerns latent attributes. The endorsement probability of 0.20 suggests limited recognition that latent attributes are not directly observable and require a measurement model to estimate possession. German HTA inevitably deals with latent constructs: pain, fatigue, depression, anxiety, cognition, functioning, satisfaction, health-related quality of life, treatment burden, and caregiver burden. These are not directly observable in the same way as survival time or hospital admission. They require inference from observable indicators. The scientific issue is therefore not simply whether questionnaires are used, but whether a measurement model establishes a unidimensional scale of latent attribute possession.

This is where Rasch measurement becomes unavoidable. A curriculum that teaches latent patient-reported outcomes without introducing Rasch measurement fails to provide students with the scientific framework required for latent measurement. Students need not become Rasch specialists, but they should understand that adding ordinal questionnaire responses does not create a measure. They should know that latent attributes require a model that estimates possession of the attribute from response patterns. They should understand why Rasch occupies a unique position as the framework that can satisfy representational measurement requirements for constructing latent measures. The German research and training knowledge base does not appear to make this a visible curriculum requirement.

The ninth statement receives 0.05: manifest and latent attributes require different forms of ratio measurement. This is perhaps the most demanding of the ten competencies, but it is essential for scientific HTA. Manifest attributes may support linear ratio measures where a true zero and unit structure are defined. Latent attributes require Rasch logit ratio measurement because the attribute is not directly observed. If students are not taught this distinction, they may treat all numerical outcomes as if they belonged to a common analytic category. This is precisely how curriculum inversion reproduces measurement inversion. A student learns to model, average, combine, discount, or compare numerical values without understanding that different attributes require fundamentally different forms of measurement.

The tenth statement concerns falsifiability. The endorsement probability of 0.15 indicates weak recognition in the curriculum-facing knowledge base. German HTA is certainly evidence-oriented, and its statutory benefit-assessment process is concerned with transparency, comparative evidence, and decision-making. But falsifiability is a stricter requirement. It requires prospectively specified claims that can be empirically evaluated, replicated, and refuted. A therapy-impact claim should specify the target attribute, population, comparator, measurement method, time horizon, success criterion, and the evidence that would count against the claim. This is different from a modelled

projection or a general claim of benefit. If students are taught to appraise evidence but not to formulate falsifiable therapy-impact claims, then they remain within evidence administration rather than scientific claim testing.

Taken together, the results indicate curriculum inversion in the German HTA research and training knowledge base. The pattern is not one of complete absence of scientific concern. Germany's HTA system is highly developed, evidence-oriented, and institutionally mature. The issue is more specific and more fundamental: the knowledge base does not visibly require students and practitioners to master the principles of representational measurement before they encounter HTA methods. The curriculum therefore appears to privilege benefit assessment, evidence appraisal, decision-making, and methodological procedure over the prior scientific question of what constitutes measurement.

This matters because professional competence in HTA cannot be defined solely by familiarity with German benefit assessment, IQWiG methods, G-BA procedures, EUnetHTA requirements, real-world evidence, or economic evaluation. These may be important professional skills, but they do not by themselves establish scientific competence. A graduate may know how to prepare a dossier, interpret an assessment, perform evidence synthesis, evaluate comparative effectiveness, or populate a model while remaining unable to explain why measurement must precede arithmetic. Such a graduate may be technically proficient but scientifically underprepared.

The question, therefore, is not whether German HTA training produces capable analysts. It does. The question is whether it certifies graduates as competent in scientific HTA. Would a graduate be expected to explain why multiplication requires ratio measurement? Would they be expected to distinguish ordinal scores from ratio measures? Would they be expected to explain why utilities cannot be assumed to possess measurement properties simply because they are numerical? Would they be expected to distinguish a manifest attribute from a latent attribute? Would they be expected to understand why Rasch measurement is required for latent therapy-impact claims? Would they be expected to formulate a falsifiable claim rather than a modelled projection? If these questions are absent from the curriculum, then curriculum inversion has occurred.

The implications extend beyond Germany. Germany is not simply a national HTA system operating in isolation. It is a major participant in the development of European HTA standards. IQWiG and G-BA have participated in EU methodological preparation through EUnetHTA21, and German requirements have been discussed in relation to the emerging joint clinical assessment framework. If German HTA training reproduces measurement inversion, then the effect is not limited to German students. It may influence the European knowledge environment through methods, dossiers, guidance, joint assessments, publications, professional expectations, and university teaching.

This is why curriculum inversion is more serious than a missing module. It affects the reproduction of an entire methodological culture. Students trained within an inverted curriculum become researchers, reviewers, agency staff, consultants, journal referees, and teachers. They then reproduce the same assumptions in professional practice. Guidance shapes curricula. Curricula shape professional competence. Professional competence shapes submissions, assessments, publications, and peer review. The result is a self-reinforcing knowledge system in which accepted

methods are taught because they are accepted, not because their measurement foundations have been established.

## **CONCLUSION**

The German HTA research and training knowledge base faces a straightforward test. It should ask whether every graduate entering HTA practice can answer the ten curriculum questions. Can they define the target attribute? Can they explain why therapy assessment begins with the attribute? Can they distinguish nominal, ordinal, interval, and ratio scales? Can they explain why measurement status must precede quantitative claims? Can they describe the role of representational measurement? Can they explain unidimensionality? Can they distinguish manifest and latent attributes? Can they explain why latent attributes require a measurement model? Can they distinguish forms of ratio measurement? Can they explain falsifiability?

If the answer is no, the problem is not merely pedagogical. It is professional. German HTA education would be certifying competence in established procedures without certifying competence in the scientific principles required to evaluate those procedures. That is the defining feature of curriculum inversion.

The conclusion is direct. The German HTA research and training knowledge base appears strong in empirical evidence appraisal, benefit assessment, public decision-making, and institutional procedure. It appears weak in the measurement foundations required for scientific HTA. This does not diminish the sophistication of German HTA. Rather, it identifies the scientific foundation that must be restored if HTA is to claim standing as a quantitative discipline. Germany should therefore undertake an explicit national assessment of curriculum inversion across its HTA agencies, universities, research centers, professional training programs, journals, and dossier expectations. The purpose would not be to impose a predetermined reform but to determine whether the scientific competencies required for lawful quantitative claims are present.

Only after such an assessment can Germany decide how best to proceed. If the findings are confirmed, then the question will no longer be whether reconstruction is necessary but how it should be undertaken within Germany's own legal, academic, and institutional traditions. The first step is recognition: scientific HTA must begin with measurement, not arithmetic; with attributes, not models; with representational measurement, not numerical convention; and with evaluable, replicable, and falsifiable therapy-impact claims rather than continued reliance on methodological inheritance.

## **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

---

<sup>i</sup> Maimon Research LLC. Interrogations in HTA. [www.maimonresearch.com](http://www.maimonresearch.com)

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

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

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

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