A Framework for Assessed Cognitive Abilities in Entrance Examinations

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Manav Sivaram

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2024

Supervisor: G Nagarjuna

Certificate

This is to certify that this dissertation entitled **A Framework to Analyse Assessed Cognitive Abilities in Entrance Examinations** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Manav Sivaram at Indian Institute of Science Education and Research under the supervision of G Nagarjuna, Visiting Faculty, Department of Science Education, during the academic year 2023-2024.

Committee:

G Nagarjuna

Pooja Sancheti

This thesis is dedicated to

Kota ke bacche

Declaration

I hereby declare that the matter embodied in the report entitled **A Framework to Analyse Assessed Cognitive Abilities in Entrance Examinations** are the results of the work carried out by me at the Department of Science Education , Indian Institute of Science Education & Research (IISER) Pune, under the supervision of Prof. G Nagarjuna, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Manav Sivaram

20191142

Table of Contents

Acknowledgements

I would not like to believe that I would’ve penned this thesis if I did not have some amazing humans around me.

I would like to thank GN for being the most patient supervisor. Beginning with conceptualising the research questions, he motivated me to do the project at each step when I felt we are at a dead-end, or doubted my ability to drive this project. Looking back, I would have missed out a great deal not working with him: the values he holds dear, his approach to problems, and his mentorship. I cannot thank him enough for all the oddly-timed meetings.

The analysis part of wouldn’t have been possible without Vedant’s excel skills. His assistance from running errands to being technician almost deserves a co-authorship.

I am grateful to Sabareesh for all the discussions and inputs to the project, which helped to fine tune many parts of the framework proposed here, for being a big brother who looked out for me, and a mentor.

I lack the vocabulary to thank Sakshi, Vedant, Arya, Abhisek, Vidisha, Spumpy, Ritvee, Akash, Lokesh, Manasvi for making me feel at home at IISER, for all the help you have given me. I will miss each of those beautifully stupid moments we shared.

Chhavi and Shalini’s mentorship and friendship when I needed it the most have played a big part in pacifying my apprehensions by holding the right mirror up at the right time. I also thank Ravi and Mo for cementing my confidence in moving forward in the field of Education.

I should never be able to put to words how much Amma, Achan, Ashoo, Ammumma, Appu, and Kichammumma mean to me. It is ultimately their belief in me that keeps me afloat and motivated. Uma maiya and Kaku need special mentions for making me a part of their homes.

It would be incomplele to close the acknowledgements without mentioning Alen, Akshay, Joe, Saurav, and Lynn for being an inseparable part of my self.

Abstract

The study aims to understand the cognitive abilities that would constitute an ideal aspirant for Higher Education in STEM. Building up on the idea that aptitude can be understood in terms of abilities, we analyse the question papers of two entrance examinations, JEE-Advanced and the IISER Aptitude Test, for the cognitive abilities they test for. Using prerequisite analysis of some major cognitive abilities that higher education would confer, we develop a framework of cognitive abilities. The framework is used to code the question items of the question papers and is analysed. We take into consideration the effects of previous instruction which affects how an aspirant would answer a question. The study points to an over-representation of content-oriented testing in the entrance examinations, calling on for recollection or procedural abilities. We also point out the limitations of the study, and further analysis that we deem will be informative.

Introduction

Entrance Examinations and Selection Procedures

Entrance Examinations have been ubiquitously used as screening tests to Higher Education in India. Perhaps they are more frequently used in India than elsewhere. They have been used as screening tests to decide students' eligibility to get admitted to an institute of preference. The most common method is to conduct the examination, and rank the students on the basis of their score. The eligibility to get admitted into an institution of the aspirant's preference is determined by their rank and a general criterion of percentage score in the higher secondary examination. This pattern of testing is referred to as "norm-referenced testing" (Turnbull, 1989). The number of entrance examinations has changed over the years. There have been attempts to lessen the number of entrance examinations to streamline admissions to various institutions. The last decade saw many of these attempts. The Indian Institutes and National Institutes have also positioned themselves as institutes of eminence. These factors, along with socio-psychological factors, have set up and partially legitimised an atmosphere of competition. A tightly-linked concept is that of "merit". In the past years, there have been attempts at understanding the nature of merit. These studies view merit from a sociological perspective. They reveal patterns of historical injustice on communities and point out structural limitations which do not remedy, and possibly reinforce the inequities. In jist, these studies concern the non-levelness of the playing field (Deshpande, 2006; Madan, 2007). Although these studies are insightful, they treat the instrument of selection, the procedure, and the question paper as black-boxes. However, it is imperative to look at the instrument, as it could inherently bias the selection process, reinforcing the social inequities. The necessity for change in the pattern of questions has been pointed out before (e.g., Mohanan, 2010).

Each entrance examination is a standardised test. A test is considered standardised if 1) all examinees attempt the same question and 2) if the scoring method is pre-determined. The modern standardised test format usually comprises Multiple Choice Questions (MCQ). In many formats, each question has one “correct” answer yielding a fixed positive score, and any “wrong” response (often referred to as distractors) might be scored with 0 or negatively. Such standardisation is deemed necessary to avoid any bias that may otherwise affect an aspirant’s chance of having a fair evaluation. The essential feature of such MCQ, and their variants (e.g., Multiple-Select Questions) is the quality of the distractors. A good distractor would allow pointing out a misconception in the student’s understanding of knowledge, or an error in applying procedures due to such misconception. In this respect, MCQs are good instruments for probing misconceptions. It needs to be noted that this paradigm of testing has in it implicit that a student is expected to know the “correct” conception. The MCQ-type questions format of standardised testing has also been criticised on the grounds that it leads to the mechanisation of learning (Punjabi, 2020).

The Indian Institutes of Technology (IITs) and the Indian Institutes of Science Education and Research (IISERs) are seen as prestigious institutions. Although in the current view of Science, Technology, Engineering and Mathematics as broadly united (STEM), a distinction between IITs and IISERs might not be meaningful. Further, these institutions also conduct research and teach other disciplines than those of natural sciences. However, there could be distinctions drawn between the learning outcomes of a student enrolled in the two institutions, which are often policy-level distinctions. The existence of the policy-level distinctions could be justified at the level of expected research output or at the level of structural limitations, e.g., the maximum capacity of the institute to cater to the pedagogical needs of students in a year at a given level. This may also justify the use of different modes of selection procedures to these institutions. For simplicity, if we assume that each institution has an exclusive mode of selection, then one would expect the selection procedure to be distinct, to whatever degree, to reflect the distinction at the level of Learning Outcomes. Such an expectation is motivated by the constructivist paradigm of learning. As the entrance examinations are a major component in the selection procedure, it is important to understand the role they play via a characterisation of their affordances. Perceptions of entrance examinations have gaps between aspirants and the professors in the institutions they select for. These arise from a mismatch of what is perceived as quality and useful education (Punjabi, 2020).

Mohanan (2010) points out that the question paper predominantly checks "(a) an understanding of a body of established knowledge; (b) the ability to apply that knowledge to standard textbook problems and situations". Mohanan suggests a value system (non-exhaustive) for the IITs that should be reflected in the entrance examinations. We agree his argument that the aspirant "should have the potential to acquire these attributes through the education that the [the institute] provide[s]." This notion of probing the potential is denoted, in the literature, by the concept of *aptitude*. It is unclear, in the absence of policy documents, if the current examinations are aptitude tests or mere screening tests. The distinction between the two could also relate to the conceptual distinction between *merit* and *aptitude*.

Taking inspiration from the value system proposed by Mohanan (2010), we note the need for a framework to describe aptitude based on the cognitive abilities assessed in the entrance examinations. However, there is a need to delineate what constitutes an ideal aspirant to the BSMS or BTech program. This is important in the light of the above discussion on aptitude. Once such an agent is constructed, the examination instruments can be compared to see what are abilities are tested by them.

Research Questions

The previous discussion brings us to two questions:

1. What does it mean to say an aspirant has aptitude for the BSMS or BTech programs?
2. Does the entrance examination do a good job of distinguishing those aspirants with this aptitude?

The BS-MS program at IISER began as an attempt to incorporate hands-on research at the undergraduate basic sciences program. The curriculum structure of IISER needs to, then, be evaluated in terms of its deviation from those undergraduate programs of other universities. If the deviation is characterised as a focus on research, the element should be pervasive in the curriculum. The program can then be characterised as a set of learning outcomes. Such a set should be listed in its Curriculum Framework. In an agent-oriented approach, these learning outcomes can be modelled as abilities that the agent will possess at the end of the course[1]. The abilities-based approach allows one to probe the success or failure of the instruction through the behaviour of the agent, that is how the agent interacts with a situation. In the case of the BS-MS program, the Learning Outcomes could broadly said to be the ability to engage in the practice of science (to be explicated in a following chapter).

Aptitude refers to a concept that intends to capture the suitability of an agent to acquire a certain ability. Russo (2011) cites Cronbach and Snow's definition of aptitude: "psychologically, aptitude is whatever makes a person ready to learn in a particular situation". Russo points out that such a definition makes selecting suitable aspirants more than just about structural limitations (such as maximum number of people who can be accommodated), but also about ethical correctness as the ones who are not selected as deemed that they do not have the aptitude. In general, there seems to be consensus that aptitude needs to be understood in relation to the practice for which the agent is selected (Chaudhuri et al., 2021; Mohanan, 2010; Russo, 2011).

[1] These abilities have also been called "competencies" in literature

That is, in the case of IISER, the aptitude of an aspirant needs to be understood in relation to the practice of science. We take this position as axiomatic for the current study.

A concrete description of an ideal aspirant is currently lacking. Given that aptitude needs to be understood as abilities that will allow students to learn the abilities to become a scientist, a description of the practice of science in terms of abilities is necessary. The characterisation of scientific reasoning as model-based reasoning is an apt framework, as it has implicit an agent-based approach (Rost & Knuuttila, 2022). We propose to create a framework for abilities involved in scientific practice and identify cognitive dependencies between them. This is based on the constructivist idea of learning as building up on what is already known (Sawyer, 2006). The ideal aspirant will then be constructed as an agent with a set of these abilities identified through the dependency analysis. The entrance examination, in line with the aptitude assumption, should be an instrument to probe these abilities that constitute the ideal aspirant.

The two Research Questions can then be formulated thus:

1. Assuming that the set of abilities that will be developed in the students through the undergraduate program is well defined, what is the set of pre-requisite abilities that an aspirant is expected to have?
2. What set of abilities does the question papers of the IAT and JEE require from the aspirants in order to be solved successfully and what does this success mean in terms of the pre-requisite abilities defined in response to 1?

A First Look at Question Papers

The items of the question papers of the IISER Aptitude Test of 2023 and 2022 were tagged for the following question: which abilities are needed answer this question correctly? This exercise was done to obtain a preliminary understanding of the nature of the question paper and to understand which abilities are asked of the examinee. The following list was obtained:

a. Recollection - of information, definition, formulae, equations

b. Procedures - Problem-solving recipes, sign manipulations, arithmetic, optimisation

c. Visualisation - representations, structures

d. Heuristics

e. Reasoning - synthesising information, logical calculus

f. Conceptual Understanding

Although useful for a primary understanding, these traits are not well-defined. Some of the above be classified as or decomposed into behaviours or learning traits more easily than others. Items d, e, and f in the above list are more suitably classified as learning traits than behaviours. However, their behavioural correlates, and the relation between them need to be well-defined. An early attempt of doing this, which is currently used, is Bloom's Taxonomy of educational goals.

Bloom's Taxonomy

Bloom's Taxonomy refers to a taxonomy of "educational goals/objectives" proposed by Benjamin S. Bloom and others in 1956 to facilitate better communication between examiners and educationists. The taxonomy assumes that "educational objectives stated in behavioral form have their counterparts in the behavior of individuals". It then provides a classification of the various learning objectives that were in use at the time, i.e., it classified the educational goals that were already identified. The taxonomy has been widely used in lesson planning and curriculum design, as it provided an outcome-oriented approach and enabled discussions that used standardised terms. The taxonomy identifies six classes of learning objectives: (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5) Synthesis, and (6) Evaluation. Each class might have sub-classes that identify the constitutive behaviours. For example, Knowledge has distinct sub-classes for remembering specific information, information of conventions, information of theories. The class of Comprehension identifies three distinct subclasses: translation, interpretation, and extrapolation (Bloom et al., 1956). It also needs to be noted that the framework is discipline-neutral.

The major classes are organised in a nested manner. A class is considered as a composite of all preceding classes, meaning the behaviours in the preceding classes are part of the behaviours in the class. The taxonomy has, then, a A-AB-ABC ordered structure (Bloom et al., 1956). If this ordering is to be accepted, it implies that a preceding class not having any behaviour of a succeeding class. Such a proposition is not practically valid even within the large paradigm of the framework (Soozandehfar & Adeli, 2016). Evaluation and Synthesis cannot be seen as ordered in the suggested manner as the two classes are required to check the validity of the other. In the context Analysis, the authors admit that “no entirely clear lines can be drawn between analysis and comprehension at one end or between analysis and evaluation at the other.” The distinction is justified as “comprehension deals with content [...], analysis with both content and form”, and that “one who is skillful in the analysis [...] may evaluate it badly” (Bloom et al., 1956; Soozandehfar & Adeli, 2016). Such distinctions do not hold in light of the current philosophy of science.

Bloom’s Taxonomy is set largely in the behaviourist paradigm although there is an influence of the constructivst theory of learning (Bertucio, 2017; Patwardhan et al., 2022; Soozandehfar & Adeli, 2016). Since 1950s, advancements in cognitive science and educational research have improved the understanding of how humans learn. The taxonomy is not current with such advancements. The taxonomy distinguishes between concepts and the actions they enable. For example, the Knowledge class has as subclasses "information of procedure", but the behaviour of application of this information is identified separately and further in the ordered manner: it is meaningful that 'one many know the procedure but may not know how to apply', however, 'one knows how to apply the procedure, but does not have the information about the procedure' is meaningless. The contradiction is apparent if one is to consider the definition of the class each of the behaviour belongs to. Further, Soozandehfar & Adeli point out that there are large inconsistencies in how educators interpret and implement the objectives.

The ordering also implies an order of cognitive complexity of the behaviour, classifying some classes as lower-order and others as higher-order. Bloom et al. (1956) point to consciousness or awareness as a correlate for this complexity. Soozandehfar & Adeli (2016) argue that "being referred to as lower order skills does not make knowledge or comprehension any less important, rather they are arguably the most important cognitive skills because knowledge of and comprehension of a subject is vital in advancing up the levels of the taxonomy."

Anderson and Krathwohl presented a revision of the Taxonomy in 2001. The Revised Taxonomy, however, only attempted to make changes in the details without challenging the underlying theoretical assumptions (Soozandehfar & Adeli, 2016). These lacunes can be solved only if a framework is created from different theoretical assumptions derived from philosophical considerations and perspectives on learning.

Theoretical Perspective

A framework of abilities relevant to education needs to correspond to the practice it aims to develop in an agent. We take the aim of education to impart tenets of scientific reasoning. The value system proposed by Patwardhan et al. (2022) provides is pitched at the level of Higher Education. However, it can be argued that a similar list of abilities can be used at the School Education level. The difference is that of expertise, which marks quantitative and qualitative differences in knowledge structures. However, expertise can only be understood once the underlying knowledge structure is well understood (Sawyer, 2006). A current opinion in the philosophy of science views scientific reasoning as model-based reasoning. This conception of scientific reasoning aimed to bridge between "mental models" and "external models", what Nerssessian called *cognitive partnership* (Nersessian, 1999b; Sawyer, 2006). An exploration of model-based reasoning allows to meaningfully identify abilities relevant to STEM Education.

Data from previous research bring to light the gap in perception of aspirants, and other stakeholders about the quality of questions asked in the examinations. Punjabi (2020) points describe the effect of the examination design on student learning perceptions. The question papers employed in the entrance examinations need to be characterised using the abilities identified. An aptitude test should be able to discriminate between high and low-aptitude aspirants.

Science as Model-Based Reasoning

The current discourse in the Philosophy of Science relies heavily on the conception of scientific reasoning as model-based reasoning.

Morrison and Morgan (1999) suggest that models mediate between theory and the world. They are seen as tools that help the investigation of a phenomenon. The nature of the scientific investigation is that of providing an explanan for the explanandum, subject to certain conditions imposed on the nature of the explanan. These conditions constitute the epistemology of science. The notion of theory can be characterised as such an explanan for those observations' generalisation in a domain of phenomena. However, it has been noted that theories are seldom capable of generating the phenomena - a criticism of the syntactic view. On the other hand, it is also observed that theoretical principles do not directly apply to specific phenomena. That is, the explanan the theory offers does not account for the specifics of the phenomenon directly. They need to be contextualised for the phenomenon, i.e., the theory needs to be mediated to fit the data (a notion explored in a later section). It is in this context that Morrison and Morgan (ibid.) propose to view models as epistemic tools that are partially independent of theory. Although the phenomenon has its specifics, not all details are relevant for a certain explanandum. In fact, a complete description of phenomena might be debilitating to the explanation, if not impossible per se. In the construction of the explanan, then some details are not considered - referred to as abstraction. The model, an abstraction, is partially independent of the phenomenon, or "the real world". This perspective on models enable to see models as mediators between theory and the real world. Knuuttilla (2011) further argues to view models as epistemic artefacts that allow for explanations, moving away from the theory-oriented discourse. It also remains a point of contention if theory and model are distinct objects.

It is imperative to explore what accords the view of epistemic tools or instruments to models. They are instruments in the sense, as noted above, that they allows the user to investigate a phenomenon. The nature of explanation is more general - that of generativity. It should be able to adequately generate an explanan, where adequacy is determined by the explanandum. Further, a model should be able to generate possibilities which could be empirically validated. As a result of abstraction, models also generate unactualised possibilities. Hesse refer to these as analogies - a relation between the model and the phenomenon (Morgan & Morrison, 1999). As the model is also influenced by the theory, models also allow the user to probe further into the theory. In the earlier discussions of what allows such learning (i.e., about the phenomenon) by using models, suggestions were made for structural relationships between the model and the real-world. However, using the same model for multiple unrelated phenomena (e.g., the Lotka-Volterra model) indicates that they're more a feature of the abstraction. The model does not replicate the phenomenon but acts as a representative of the system to the extent that the explanandum demands. This also implies that it is not the "representational power", but the "explanatory power" of the model that is necessary. If not structural similarity relations, what would explain the explanatory power of models? Boumans (1999) and Rost & Knuuttila (2022) argue that this explanatory power is derived from constructing the model themselves.

The explanatory power of the model is tightly linked to the the importance placed on the explanandum in the previous paragraph. There is a notion of *models for* in the literature. It characterises models as constructed for the explanation of a phenomenon. The complimentary notion is that of *models of*, which concern using models for pedagogical purposes. The construction is of special importance as much of the model's functionality is derived thence - "the purposeful design". The structural relations in a model are usually constrained by the question it seeks to answer (the explanandum). The explanandum itself is "motivated by theoretical and/or empirical considerations" (Rost & Knuuttila, 2022). The abstractions made about the phenomenon are introduced into the model via representational tools. The representational tools could be of different materiality (determined by their media) and often could have implications for the model's affordances. The representational tools allow for the model to be manipulated during its construction and application. It is the manipulation that allows learning. The determination of the explanandum and the choice of representational tools both seem to be understood currently as a creative enterprise. However, they are influenced by past knowledge (theoretical, empirical and of representational tools) which “build in” the justification to the model. Nersessian (1999) argues that the constrained construction facilitating learning helps to identify the context of justification with the context of discovery.

(Nersessian, 1999a) identification of the context of discovery and the context of justification points to an important feature of the constructivist paradigm: learning is discovery. In the light that scientific reasoning (and practice) can be understood as model-based reasoning, one could argue that science education should reflect this position. Rost & Knuuttila (2022) argue that science education should focus on the modal dimension of modeling. This artifactual perspective on models also allows an agent-oriented approach to understanding how science is practiced, and, therefore can be translated into the educational paradigm.

Constructing the Framework

A Value System for Higher Education

Patwardhan et al., (2022) propose, as the aim of Higher Education, the development of Academic Intelligence. In a succinct form, Academic Intelligence includes the following capacities:

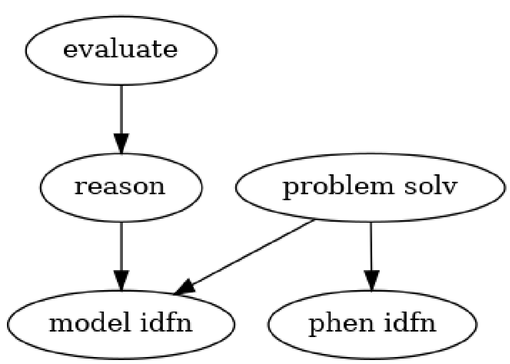
- Capacity to access relevant information and use them as sources of documented knowledge for learning independently
- Capacity to evaluate information as an educated non-specialist
    - Accept, reject or set aside for review
- Capacity to reason to reach conclusions and derive inferences
    - Ability to point out logical contradictions
    - Ability to make rational and ethical decisions
- Capacity to spell out the steps of reasoning
- Capacity to evaluate reasoning and argumentation
- Understanding of arguments in favour and against core ideas in academic knowledge
- Capacity to look out for counterexamples
- Capacity to gather data to test empirical claim
- Capacity for critical thinking, inquiry and integration
- Ability to engage intelligently and effectively in debate on public issues
    - awareness of one's own biases
    - Evaluation of policy, action, products, etc
- Appreciate the world of ideas
    - Appreciate the aesthetics of mathematical/scientific process
    - Awareness and understanding of the tentativeness of human knowledge
- Capacity to communicate ideas clearly and precisely
    - Willingness to exchange ideas and beliefs

Using the value system as a starting point, we propose four major abilities an ideal IISER aspirant should possess: **model construction, knowledge evaluation, experimentation,** and **problem-solving**. We do not suggest these four abilities are independent of each. Practice and progress of science rely on all four of these abilities. This account is explored in the chapter on Model-based Reasoning.

Pre-Requisite Analysis and Dependencies

The constructivist paradigm motivates the conception of dependencies between the various abiilities. Such dependencies could be derived from definiendum-definen relationships. The dependencies could in turn have other dependencies. This can be represented as a directed graph (Nagarjuna et al., 2010). Each ability can be represented as a node, and an edge would represent the dependency relation. The dependency relation, here called a pre-requisite relation, shows a cognitive dependency of an ability on another while learning. The pre-requisite ability is a called a "critical dependency". The relationship is, but, not that of composition. If an ability has three pre-requisites, the ability cannot be described completely as a composed of the pre-requisites, as there could be behaviours that enable the execution of the ability. Here, we mark a difference from Bloom's Taxonomy, which attempted to define the actions involved. Following Nagarjuna et al., (2010), such a dependency graph "can be classified as a species of causal dependency in the context of semantics."

To provide an example, to be able to evaluate the explanatory power of a model, an agent would need to know how to reason out predictions, i.e., construct an argument, using the model. To reason through, one needs to be able to identify the premises of the model and should understand the structure of the model. Likewise, an agent will be able to do problem-solving only if it will be able to identify the phenomenon involved, and identify a suitable model. This example can be represented as the following graph.



To identify the pre-requisite abilities of any ability, we use the definiendum-definien relationship. The definien will point out the critical dependencies of the ability. Further, any other cognitive dependency is identified through pedagogical perspectives. The pedagogical perspectives also constrain some of the dependency relationships, sometimes avoiding circular dependencies (those related in the form A -> B and B -> A) which might be present in the practice of science. The next chapter describes the definitions of each ability, and identifies their dependencies.

Each node in the graph represents an ability. Each edge indicates the relation “has the pre-requisite”. Following this, each edge denotes: *<ability 1> has the pre-requisite <ability 2>*. For example, *the ability to identify the model has the pre-requisite ability to identify the variables*. The inverse of the edge represents a necessary condition: *<ability 2> is a necessary condition for <ability 1>*.

Iterative Design

The distinction made between the practice of science and the pedagogy of science could sometime generate structures whose validity needs to be checked. These could be either ability whose identification as distinct from another ability might be a hair-splitting difference in the view of cognitive science, or could be dependencies that might not be meaningful. These structures are included or excluded by considering the meaningfulness of the natural language representation of the edges it is a part of. This method is called iterative design. Each iteration is an improvement over the previous (Wong & Park, 2010).

The graph is visualised using the Graphviz software, an open-source graph visualisation tool.

The Framework

We note that the mode of education in the Indian context is largely knowledge-oriented. This orientation requires an explication of the ability-oriented framework with the content-oriented framework. In addition, the definitions are operationalised at the level at which the test is pitched without losing sight of the educational system in its practiced form. These considerations also lead us to draw some distinctions that might seem hair-splitting in philosophical considerations. However they are essential in the context of the intended analysis and are clarified where made.

We use the propositional medium to demonstrate the relationship between ability-oriented and content-oriented framework. In this view, a model can be defined as:

> A system of coherent and logically consistent propositions that explain a phenomenon

This representation of models implicitly places the model in conjunction with the phenomenon. There is no contradiction here with the notion that models are autonomous artifacts, as proposed in a previous chapter, as explanation is enabled. The *propositional characterisation* of models helps to demonstrate the relevant abilities concretely. This is not to suggest lacunae in other representational media, but only to facilitate the description of the logic involved. As evident in the ensuing discussion, other semiotic devices (e.g., mathematical symbols, graphs, and diagrams) are used, where convenient and appropriate, complementary to the propositional medium.

Defining the Abilities

The four major abilities proposed here, which describe the practice of science are: Problem-solving, Experimentation, Knowledge Evaluation, and Model Construction. In this section we define these abilities and identify their dependencies. Some dependencies are analyse further.

Problem-solving

As the words suggest, problem-solving can be characterised as finding an optimal solution to an identified problem. In the model-based research literature, problem-solving is viewed as the

identification of an explanandum (the problem), the identification of the phenomenon, and the application (including via construction) of an appropriate model to provide an explanan (the Mathematics solution). Hence, two abilities involved in problem-solving are identified: a) the ability to identify the phenomenon and b) the ability to identify an appropriate model that would provide the solution. We propose a third dependency which concerns the application of the model to the phenomenon and the processes involved in deriving the solution: c) the ability to carry out procedures.

The necessity to view the ability to carry out procedures separately is due to the presence of question items that do not require the examinee to reflect on either the phenomenon or the model. In the sciences, i.e., the physics, chemistry and biology sections, such questions could have diagnosed other relevant abilities[2]. However, a question type might have been used as an instance of the model application for pedagogical purposes. Such questions usually restrict themselves to deriving the numerical values of an involved variable, given the values of other

[2] The case of Mathematics is discussed separately.

involved variables, usually in an indirect manner. It could be argued that the indirect presentation of values prompts reasoning abilities; some examinees would indeed build the model, identify the boundary conditions imposed on the problem, and find the solution via reasoning. This argument is not rejected. However, if a similar question has been repeated over previous iterations or if the question type is typically used for pedagogical purposes, the examinee could passively derive the solution. In such a case, the question fails to be diagnostic of the ability it tests.

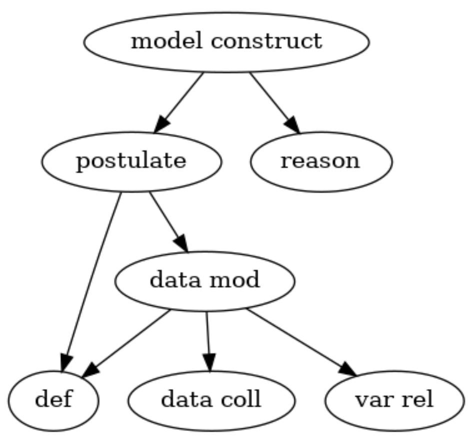
In a case where such a passive response (template-driven) is less likely, the examinee would have to actively engage with the question in either identifying the phenomenon, identifying a suitable model, or both. Further, there could be multiple solutions possible, however, the identification of the optimal solution would require the examinee to compare different models (explored under Model Comparison). The content-oriented perspective would here point to the variety of procedural methods that are diagnosed. While the procedural methods are of importance, a passive application might not be useful while diagnosing aptitude, as the routinised procedures are seldom directly transferable to other contexts.

Model Construction

Model construction is the ability to construct a model to provide an explanan for an explanandum. Epistemological restrictions are placed during the construction of the model. One is that the model should be generative. It should be able to generate the phenomenon for which it is constructed, and also be able to generate other related phenomena. This has also been called modal modelling (Rost & Knuuttila, 2022).

Modelling is a creative process. Models are constructed with the purpose set by the explanandum. The phenomenon of interest is identified through the identification of the ‘observation generalisations’, referred to as idealisation. Further, postulates are constructed that would form the structure of the model. The postulation step involves proposing both objects and relationships between them. These structural elements also need to be well-defined in the context. They are not necessarily representative of the phenomenon through correspondence rules. The model is required to be logically consistent and allow itself to deduce testable predictions. The model is often tested against other related phenomena, failing which the model is updated.

We identify the following dependencies: (1) the ability to postulate, and (2) the ability to reason. The ability to postulate has dependecy on the ability to draw observation generalisations, the ability to collect relevant data, and the ability to define concepts and relationships between them. Postulates are derived abductively from the observation generalisations, as the generalisations describe the behaviour of the system. These generalisations constrain the model’s structure. The generalisations themselves are derived from available data, which are based on the definitions assumed. There is, arguably, a cyclicity in the practice, but we propose the dependencies from a pedagogical perspective. The dependencies are represented in the following graph.



Knowledge Evaluation

A knowledge claim about a phenomenon can be evaluated from two perspectives: (1) in relation to a given model, if it is consistent within a given model, and (2) in relation to multiple models, a comparison of where the claim is best accommodated. To check the logical consistency of a claim within a model, deduce the claim from the premises of a model. The deduction will rely on the model's structure with reference to which the phenomenon is represented. A phenomenon could have multiple models, each with a different explanatory power. A knowledge claim might be either accommodated or rejected by a model. Two considerations are important in analysing such a mismatch: (a) empirical soundness of the claim and (b) modality of the model. If the claim is experimentally sound, that is, one finds evidence for its "truth", the claim needs to be accommodated by a model of the phenomenon. The presence of multiple models for a phenomenon could lead to different treatments of the claim. Morgan & Morrison (1999)use the example of models of atomic nuclei to demonstrate this. Some observations are explained exclusively by one model, and other observations by another. In such a situation, an agent should be able to compare between the models to evaluate the knowledge claim. Such a comparison can be made based on the assumptions the models make and their explanatory powers.

We identify the following dependencies for the ability to evaluate a knowledge claim: (1) the ability to check the logical consistency of a claim within a model and (2) the ability to compare models to decide the value of the claim. Both of these abilities require the ability to reason using an appropriate logic. Reasoning requires identifying the premises of a model and then using appropriate logic (deductive or otherwise) to derive a conclusion.

Experimentation

The ability to experiment is central to the practice of science. Experimentation may be done to check the soundness of a claim or validate a model. In either case, an agent should be able to design experiments. The design of experiments depends on two factors: identifying the relationships between different variables and the technical skills required to carry out the experiment. Identifying the relevant relationships between variables is motivated by the conceptual relationships structured in a model. The technical skills include the ability to measure the variables the interest, and also the ability to regulate the other variables.

The distinction drawn between variable and concept in the previous paragraph is subtle. We accept Knuuttila's argument that a model cannot be required to represent a phenomenon but only provide as an epistemic tool. This characteristic can be enabled only by a distinction of what the model proposes and what may be observed. A variable could loosely be described as a function of the concepts involved in a model. The variable is measurable, through which a model enables empirical validation.

Descriptions of other abilities

Model Identification

The ability to identify a suitable model is enabled by the identification of the phenomenon. We do not propose to distinguish between mere knowledge of a model and understanding the structure of the model. A certain phenomenon might be explained by more than one model, each with differing explanatory power. In such a case, identifying the model that would provide the best explanan needs to be identified. Such a model might be derived from a theoretical perspectives or be phenomenological, or borrowed from previous knowledge; a distinction between the two is unnecessary as only suitability matters. Suitability is decided by the explanandum, and to the extent that the explanandum is addressed, a distinction between “constructing” and “applying” models is irrelevant.

The choice of the model, as is contingent on the explanandum, is dependent on the identification of the relevant variables and the relationships between them. However, it might be cognitively simultaneous, a pedagogical distinction needs to be made between identifying the variables, the relations between them and the model. The identification of relations between variables would enable the choice of a model whose assumptions the relations satisfy. Some question items might imply a model and therefore only the identification of the relevant variables becomes important.

An important component of the model is the representational media. The diagrams, graphs, formulae etc. allow for the manipulation of the model. Often referred to as “graphicacy”, following literacy and numeracy, is also an ability that can be tested. Some models, especially in the case of mathematics, are identified with their representational media. Therefore we also propose the ability to use suitable representational media as a dependency of model identification. This dependency is often understated.

Phenomenon Identification

Identifying the phenomenon is important in the sense of the idealisations made. The idealisations help characterise the explanandum into a tractable form. Again, a distinction is made between model and the phenomenon identification, whereas they might be cognitively simultaneous. A pedagogical distinction is necessary to motivate the meta-modeling learning, a necessary component of model-based learning (Constantinou et al., 2019). Often, the phenomenon is explicit in a question item and does not require the agent to idealise the phenomenon. This may be due to the usage of words that denote concepts that are part of a model typically used in relation to the phenomenon.

In the context of model construction, the observation generalisations are, again, idealisations, and the dependency on identifying the phenomenon is not required to be distinguished as the phenomenon is implicit in the other dependencies.

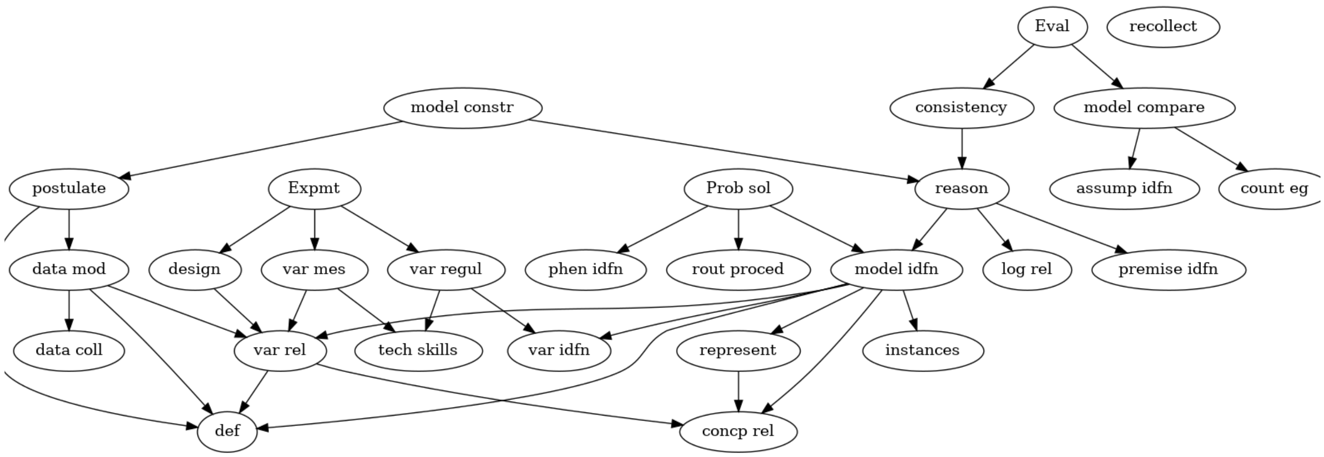
Define

The ability to define variables and concepts (including categories and relationships) is a necessary dependency for model construction. Although it is *sink* dependency (in graph-theoretic terms) in the framework proposed, it is an important ability in the practice of science. The justification for a definition relies on its position in the structure of the model. However, since structural relations rely on the definitions devised, there is a pedagogical dependency that initiates the organisation of data.

Recollect

The recollection ability is placed as a separate code in the framework. Whereas it is a dependency for each ability in a sense, it does not provide any new insight into any of the other abilities or practices. Question items in assessments often tend to only stress on the recollection abilities. Motivation is stressful for the content that needs to be conveyed. We take a different position to focus on the practice of science (explicated in Discussion).

The following graph summarises the dependencies identified in the above analysis. Each arrow is directed and indicates the dependency relationship. The graph follows the convention described in a previous section. The definitions are written down in Appendix 1. The edges are written out in Appendix 2, to check the semantic soundness of the framework.



Question Paper Description Analysis

With the framework developed, it needs to be checked whether the question papers adequately demand the abilities the ideal aspirant is expected to possess. Other considerations are the discriminability power of the question items, i.e., whether the question can diagnose one ability without involving another ability.

Data Sets

As we focus on IISERs and IITs, we use the question papers of the two entrance examinations to these institutes, the IISER Aptitude Test (IAT) and the JEE-Advanced. We use question papers for the years 2023, 2022 and 2021. The two examinations have different formats, although the syllabus is the same. The IAT includes 60 questions, 15 from each subject discipline (Physics, Chemistry, Biology and Mathematics) to be solved in 180 minutes. The questions are all of MCQ kind, with one correct answer, yielding a positive score. A wrong response awards a negative score, and no attempt is scored (0). The aspirants appear for IAT directly. The JEE-Advanced, on the other hand, is allowed only to those who qualify for the JEE-Mains examination. The JEE Advanced examination has two papers, each spanning 180 minutes. The format of both papers in a year is the same. In addition to MCQs similar to IAT, JEE-Advanced question papers include more types of questions like Multiple Select Questions (MSQs, more than one choice could be correct), integer or numerical type questions (NAT/INT, which do not provide choices), and MMQs (matching type questions). Each question format has an associated scoring pattern.

Solution manuals for each of these question papers are available online. These manuals are, typically published by individuals or firms who coach for the examinations. The solutions manuals are used to get an idea of the possible pathways to solve a question item. It is reasonable to assume that many questions will be solvable through derivations from first principles, described as problem-solving the framework proposed herein. However, the solutions presented in the solution manuals also rely on formulae that an agent could be coached to remember through repetitive practice. Each of these solutions are accepted and coded separately. It checks how many of the abilities are tested in each question paper and also if there is an over-representation of certain abilities in the cases where alternate codes are assigned.

For JEE-Advanced, the organising committee publishes analytical data regarding the conduct of the examination. The report includes the response rate (percentage of attendees) for each question, including the rates of correct and wrong responses. We are interested in identifying if there is a qualitative correlation between the response rates of a question item and the abilities tested for by it. Such a relation could provide insight into the cognitive demographic selected by the question paper.

Coding Question Items

Coding in qualitative research is categorising and labeling data to identify patterns, themes, and concepts. It involves systematically organising and interpreting textual or visual information. There are various methods and approaches to coding in qualitative research, and researchers often use a combination of them based on their research questions and data (Bizer et al., 2009; Nagarjuna et al., 2010).

We are interested in characterising the cognitive skills demanded by each question paper. In the current study, we propose to code each question item with the ability to find the correct answer. The codes used are the abilities identified in the framework created. The codes are assigned to each question using a protocol.

An outline of the protocol is sketched here:

> The solution pathway to each question is tracked. The major steps in the solution pathways are subjected to the question: "what does one need to possess/do to move ahead?". The answer to this question is sought in the list of codes.

> A convention is assumed that, if a question has been coded with a node with pre-requisites, it also has all pre-requisite codes. The question might not be able to delineate which among the prerequisites are tested and which are not.

Some questions may have multiple solution pathways. Such questions are coded with two sets of codes. The presence of such questions point to low discrimination power of the question. In such cases, we analyse if some abilities are over-represented.

Observations from Codes Analysis

The questions are classified into two categories based on the abilities tested: conventional and non-conventional. Questions that involve only recollection or procedure-oriented solutions are classified as *conventional*. If any other ability is tested, the question is classified as non-conventional. To enable meaningful interpretation, if a question item has both conventional and non-conventional codes, the conventional code is considered the alternate code.

A degree of subjectivity is involved in the coding process. In addition to the solutions prescribed in the solution manuals, it was considered if the question could be solved using abilities. In the case that such a solution was not found or, more accurately, not thought of, only one code was assigned to the question. Attempts had been, however, made to check if a different coder might assign codes differently. Any differences were considered and either accommodated or rejected.

The data from the reports on JEE-Advanced provides data for the rate of not-attempt responses to questions. The following box plot shows the distribution of question-response rates for various classes of questions based on the codes assigned.

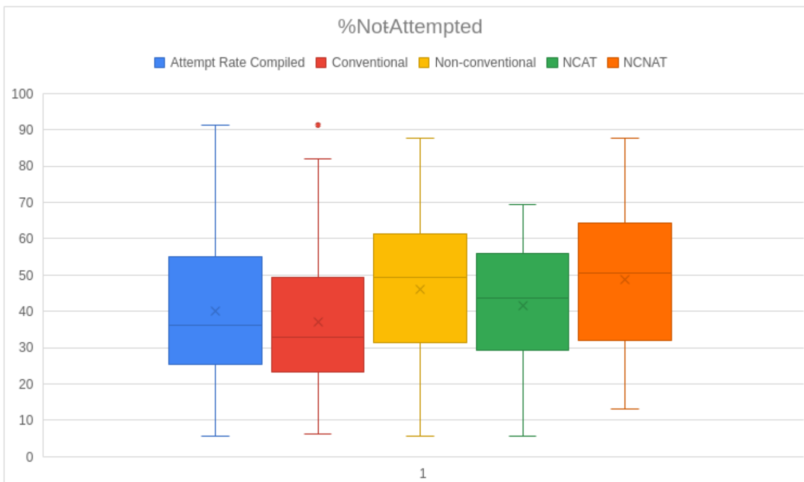


Figure shows the distribution of questions with non-attempt rates for the various classes of questions based on the codes assigned to the question items. (NCAT is non-conventional with alternate code, NCNAT is non-conventional without alternate code)

The median not-attempt rate for all questions combined is around 40%. We take this as the cut-off to decide if a question is high-attempt or low-attempt. A question is considered a high attempt if its not-attempted rate is less than 40%. We check for a correlation between the codes assigned and the attempt rate.

JEE Advanced							
Year	High Attempt (<40% not attempted)		Low Attempt		Total Question	Non Conventional with Alternate code	
	Conventional	Non-Conventional	Conventional	Non-Conventional		High Attempt	Low Attempt
2023	**69**	**22**	**4**	**5**	**100**	**6**	**0**
Physics	20	12	0	0	32	5	0
Chemistry	27	5	1	1	34	0	0
Maths	22	5	3	4	34	1	0
2022	**48**	**24**	**21**	**12**	**105**	**13**	**2**
Physics	13	9	7	3	32	4	1
Chemistry	21	10	2	3	36	0	1
Maths	14	5	12	6	37	1	1
2021	**69**	**30**	**6**	**9**	**114**	**13**	**3**
Physics	22	12	1	3	38	7	1
Chemistry	27	10	0	1	38	3	0
Maths	20	8	5	5	38	3	0

The data of the rate of wrong responses was normalised to check the rate among those who attempted.

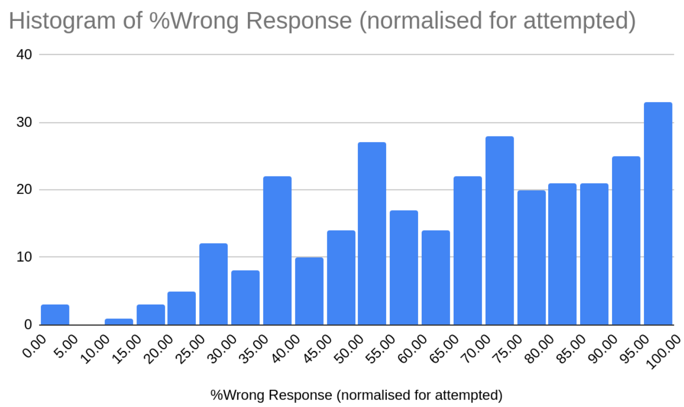
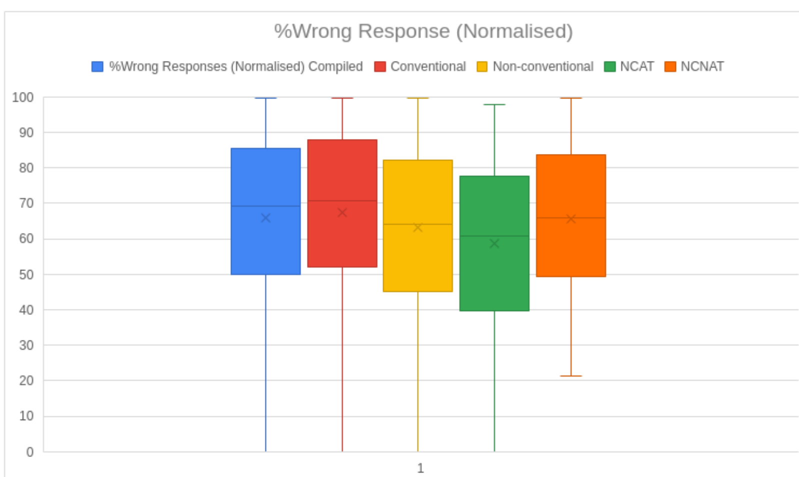


Figure shows the number of questions of certain rate of wrong response, among those attempted

For each of the class of questions, the following figure shows the distribution of questions and the rate of wrong responses (normalised)



We make the following observations from these analyses on JEE-Advanced:

1. A significant number of question items are tested only for conventional codes.
2. The median not-attempt rate among question items with a conventional code is less than that of non-conventional codes.
3. If questions of non-conventional codes have an alternate code, the median non-attempt rate is lower than those with no alternative codes.
4. Questions with high wrong response rates, among those attempted, are higher in number.
5. The median wrong response rate for questions with non-conventional codes is lower than that of questions with conventional codes.

The following table shows the number of questions with the codes for IISER Aptitude Test. Response data for IAT is not available, and hence further analysis could not be done.

IISER Aptitude Test				
Year	Conventional	Non-Conventional	Total Questions	Non-Conventional with alternate tag
2023	**34**	**26**	**60**	**12**
Physics	4	11	15	4
Chemistry	10	5	15	4
Biology	12	3	15	1
Maths	8	7	15	3
2022	**33**	**27**	**60**	**16**
Physics	5	10	15	7
Chemistry	10	5	15	4
Biology	12	3	15	1
Maths	6	9	15	4
2021	**42**	**18**	**60**	**5**
Physics	12	3	15	0
Chemistry	10	3	13	2
Biology	11	6	17	2
Maths	9	6	15	1

Discussion

We began by asking who would be considered an ideal aspirant to STEM Higher Education, specifically the BSMS and B.Tech programs. The current discourse in the Philosophy of Science and Science Education posits that STEM education be aimed at developing the practice in the students. This position is accepted in the position that views scientific reasoning as model-based reasoning. The model's function is understood in conjunction with a phenomenon, an explanandum or purpose, and the agent. This perspective lends itself to an ability-based approach, which Mohanan (2010) also suggests when proposing a value system for higher education. The selection to higher education should be of those agents with the aptitude, defined as the potential to acquire the abilities that higher education will develop. We assume that the potential to acquire these abilities can be understood as a combination of other abilities, called prerequisites. The prerequisite analysis is based on the constructivist paradigm which suggests that new abilities are built on existing abilities. The framework proposed here constructs the agent, the scientist, and the aspirant both, as a set of abilities. The distinction between the two agents is that of proficiency, or expertise. In this study, we also suggest that the aspirant's aptitude, the agent's potential to acquire the abilities that will be imparted in higher education, can be understood as a set of abilities.

With the abilities that need to be tested identified, we check if the entrance examinations, which contribute significantly to the selection process, test for these abilities. Ideally, the entrance examination would select aspirants with these abilities. We look at the question papers of JEE-Advanced and IISER Aptitude Test to analyse which abilities are tested by them. We code each question item with codes of abilities an agent needs to answer the question correctly. We have considered the significance of previous instruction to reveal some characteristics of the abilities tested by the question items. Some questions might be answered through multiple pathways, thereby receiving multiple sets of codes, which we consider two. Of particular interest is the codes “recollect” and “procedure” (here called conventional codes), because the entrance examinations and the associated coaching industry is charged with promoting “rote-learning” and “mechanisation of learning”. It would be expected that a question that has a conventional code would have less item discrimination, that is wouldn’t be able to distinguish between agents who use either set of abilities, and hence would be attempted more frequently and be solved correctly.

Observations from the Coding Process

The codes “recollect” and “procedure” are defined in the framework to refer to reproducing something from previous instruction. We consider as *content* all the descriptions of phenomena, the models and the procedures of application of the models that is included in the textbooks. We also include in *content* the exemplary questions that are provided. In the question items analysed, questions that relied solely on recollection or template-driven procedural abilities constituted around two-thirds of all questions[3]. Such questions are also significantly represented in question papers of each year of both examinations.

While relying on the same content, some questions could test for abilities that are not mere recollection or procedural application. These included reasoning abilities, identifying a suitable model, identifying the phenomenon of relevance etc. These questions were coded with such abilities by mapping solutions derived from first principles. However, if previous instruction is considered, some questions, among these, are reduced to test recollecting information or employing templates of procedures. The possibility of previous instruction is inferred from the solutions provided in the solution manuals published by “coaching institutions”. The presence of these alternate codes points to poor “item discrimination”. Item discrimination refers to the ability of a question item to discriminate between those agents who demonstrate what is intended to be tested and those agents who do not demonstrate them (Asamoah & Ocansey, 2019). We use two perspectives on item discrimination. One, at the level of the question item and the pathway to the intended response, described hitherto (call this *ability discrimination*). The other level is item discrimination at the level of attempt rates and rates of correct responses. That is, if a

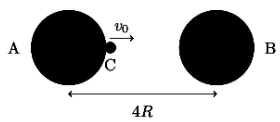
[3] We are, of course, assuming here that the questions are all distinct. To the extent that we could notice, there were seldom repetitions of the exact questions, and any repetition, even if present, wouldnot sway the general trend described.

question is lowly attempted or is more frequently responded to wrongly, the item may be said to have high discrimination (call this *response discrimination*).

A consideration that turned out to be relevant in coding the problem-solving type question items is scrutinising at what stage the problem is converted into a form that is tractable to earlier known templates.

12. Question 12

Two spherical bodies A and B each of mass M and radius R are located such that their centers are apart by a distance of $4R$ as shown in the figure. An object C of mass m is thrown from the surface of A directly towards the center of B with a speed $v_0 = 2v_{min}$, where v_{min} is the minimum speed needed for C to reach the surface of B. Given that G is the universal gravitational constant, how does the speed $v(x)$ of C change as a function of its distance x from the center of A?



CORRECT ANSWER

(a) $v(x) = \dfrac{2\sqrt{2GMR}}{x^{1/2}(4R-x)^{1/2}}$

(b) $v(x) = \dfrac{2R\sqrt{2GMR}}{x(4R-x)}$

(c) $v(x) = \dfrac{\sqrt{GMR}}{x^{1/2}(4R-x)^{1/2}}$

(d) $v(x) = \dfrac{6R^2\sqrt{2GMR}}{x^{3/2}(4R-x)^{3/2}}$

This question can be answered by identifying the relevant variables and the relation between them, which allows the routine procedures to be applied. Here, the relation could be that of energy conservation. The codes for this question are “variable relations, procedure.”

In the cases where a single step leads to a reduction, we have tagged the question item with an appropriate code, either model identification or variable identification as was deemed appropriate. In the case that such a question also had a solution pathway that relied merely on the application of a formula or procedural template, an alternative code was given. Similar considerations were applied for other codes as well.

What were considered models?

It is important at this point to consider what was considered a model for the purpose of the coding. In the following paragraphs, the subject distinctions are relevant only to the extent that they exist in the syllabus of the entrance examinations.

In the case of physics, models are mostly Newtonian. Newtonian mechanics provide sufficient theoretical concepts to describe mechanical phenomena. In light of the discussion on model-based reasoning, Newton's laws of motion could be considered as the theoretical basis for many models. The laws of conservation also serve as theoretical constraints. Models in physics include the Hooke model of spring, the Young model of solids, the Bernoulli model of fluids etc. The Coulomb model of static electricity, Ohmic models of electric current, Maxwellian model of electromagnetism, and models of behaviour of light and semiconductors are also included. In chemistry, models include the different models of atoms, (periodicity) models of the chemical nature of elements, orbital-energy models of molecular stability, thermodynamic models of chemical kinetics, Raoult model for solution interaction, etc. In biology sections, genetic models of disease patterns, mass-flow model of transportation in plants, Hardy-Weinberg and Darwinist models for natural selection are examples.[4] [5]

In Mathematics, viewing the model separately from the representational medium is difficult. In the syllabus, the different structures can be considered models. This perspective is not only in the structure in relation to the model it is employed in, but also in the relations shared between the structures. For example, both cartesian coordinates and complex numbers allow us to model 2-dimensional space.

[4] The stress on the word "model" is conscious to provoke thought of the usual presentation of these topics in the classroom.

[5] The models mentioned in the paragraph using proper nouns are only to help relate to the content that is discussed; some models might not historically been proposed by the people mentioned themselves.

Although the above paragraphs might seem to include most of the content included in the syllabus, the code "model identification" was rarely given to an item. The reason for this is to consider the consciousness behind using such a model in a given situation. In the definition of model identification, the code is given if the model is not suggested in the question itself but has to be chosen in light of the explanandum identified. The difficulty caused, pedagogically, is the identification of the model and the phenomenon in textbooks or in the classrooms, leading to a lack of epistemological awareness of the practice of science. The identification implicitly suggests, or assumes, an accurate "representation" of the phenomenon by the model, whereas these models may not consistently extend themselves to other "phenomena in the same class". Those related but unaccounted phenomena are presented as "exceptions to the rule", and not as counter-examples or limitations of the model being used. A test of "consciousness" or "intentionality" is being implied here, which is operationalised, in literature, by *meta-cognition*.

The shift of learning orientation from content to meta-cognition is towards the epistemology and practice of science. The content that is taught is usually select information from previous scientific knowledge, so they are not transferable to new problems. However, it is important to know how to solve the problem considering what constitutes a meaningful explanandum, a valid and meaningful explanan, useful evidence, what the possible counter-examples of the explanan are etc. Although it could be argued that these habits are implicit in the content that is taught, it has been demonstrated that students do not understand the practice per se. The students only carry out the procedure in a template format, although there could be a hierarchy of these templates. Knowing how to divide, knowing what to divide, knowing when to divide, and knowing why division is a good tool in a problem could be placed hierarchically. Knowing "why division works" could, arguably, be more important to be taught than training a student in dividing increasingly larger numbers beyond establishing the procedure, as the awareness of why division works enables the user to consciously employ (or refrain from) the procedure in another problem. A similar argument can be made for all other abilities. Hence, it can be argued that science education needs to focus on imparting the cultural habits of science. This does not subjugate the role of the content, which is both the substrate and an outcome of the practice.

Interpreting the Trends from Analysis

We draw the following observations from the analysed data. The data is restricted to that of JEE-Advanced.

1. A significant number of question items are tested only for conventional codes.
2. The median not-attempt rate among question items with a conventional code is less than that of non-conventional codes.
3. If questions of non-conventional codes have an alternate code, the median non-attempt rate is lower than those with no alternative codes.
4. Questions with high wrong response rates, among those attempted, are higher in number.
5. The median wrong response rate for questions with non-conventional codes is lower than that of questions with conventional codes.

In the light of the preceding discussion, it might be expected that an item that tests for abilities beyond recollection and procedural abilities should have higher item discrimination, specifically response discrimination. And conversely, high attempt rates and low wrong response rates would have more representation of conventional codes. Observations (2), and (3) align with this expectation. A question with a non-conventional code is less likely to be attempted than one which has a conventional code. If a question with a non-conventional code has an alternate code, the attempt rates are better than those without an alternate code. Observations (4) and (5) are curious. The higher frequency of questions with a high wrong response rate could indicate the questions' difficulty level. The lower median of wrong response rate in questions with non-conventional codes than those with conventional codes is also not explained. It is, however, not clear where this difficulty factor arises from. It could either be because of the presence of non-conventional codes, or because of the topic from which the questions are derived. It is noted at this point that a wrong response is, generally, marked with a negative mark. It might, hence, be more advantageous for the agent not to attempt a question than mark a wrong response. The difficulty of an item is then captured more so by the not-attempted rate than the rate of wrong response (normalised or otherwise).

These observations mark the content-orientedness of the question items, which might even contribute to the item discrimination. There is also an under-representation of non-conventional abilities in the question items. A contrast with Chowdhury et al.'s, (2023) analysis of Higher Secondary Certificate Examinations is interesting. They use Bloom's Taxonomy to classify the cognitive level used to solve question items. They describe that most questions are of the Application type. However, they do not consider if the questions are asking to replicate a template procedure from previous instructions. Needless to say, they remain in the content-oriented paradigm. The lack of consideration of how an agent would attempt a question could lead to a misidentification of the agent's cognitive features that is intended to be tested.

The higher non-attempt rates of questions with non-conventional codes also holds implications for student learning. The observations made by Punjabi (2020) provide testimony to this causal relation. A similar study, although using Bloom's Taxonomy, by Bezuidenhout & Alt (2011) shows agreeing results. This would lead us to conclude that if the question items are designed to test non-conventional abilities, it could lead to the development of these abilities in the aspirants, thereby preventing "mechanisation of learning" or "rote learning".

Limitations of the Study

The current study has its limitations in drawing more precise and informative conclusions. The limitations are the following:

1. Due to time and other resource constraints, the coding framework was not validated for consistent use. Ideally, the coding framework needs to be applied to the same question items by multiple analysers to check if the codes given are consistent to an extent.
2. The current assignment of codes might be biased and will need to be re-evaluated again, by others.
3. As the reports of the IISER Aptitude Test were not accessible, much could not be said about the question items in the examination. Although it is apparent that a significant number of questions from each year have poor ability discrimination, other dimensions of analysis could not be included.
4. The study did not consider the topic of the question as a descriptor. If such an analysis had been carried out, some trends in the different rates of response could have been resolved. Further, distractor analysis was not carried out explicitly for the question items.
5. An interesting facet of the second research question would have been how top-rankers respond to question items. For this analysis, access would be needed to the individual responses of agents. This data would also have allowed to check the reasons for the wrong response rates.

Appendix 1

Definitions of Abilities

the names in bracket are the node names in the graph

Propositional Evaluation (Eval)	Determining the value of a proposition or a set of propositions as 1) logical consistency as part of a model, or 2) the positional value of a proposition between models
Model Construction (model constr)	Constructing models through postulating from modelled data, and defining concepts
Experimentation (Expmt)	Conducting experiments to check the degree of correspondence of a proposition or a model with observations
Problem-solving (Prob sol)	Finding any or the optimal solution, generally explanan, to a problem (explanandum) within the defined boundary conditions
Comparing Models (model compare)	Comparison between models via comparison of A) soundness of premises; B) agreement of prediction with experimental observation
Identifying the assumptions (assump idfn)	Identifying the underlying assumptions of a model
Checking consistency	Checking the logical consistency of propositions within a model

(consistency)	
Reasoning (reason)	To derive the conclusions from a set of given system of premises, assuming a system of logic (explicitly stated or otherwise)
Identifying premises (premise idfn)	Identifying the major and minor premises, from which the conclusions shall be derived
Postulation (postulate)	To construct the premises of a model from modelled data
Identifying logical relations (log rel)	Logical relations between propositions: implication, negation, contrapositions, contradiction, inverse, converse etc.
Identifying instances (instances)	To identify the instances of a concept
Representing (represent)	To represent (or interpret the representation of) a model, through visual, linguistic or other means
Identifying conceptual relations (concp rel)	Identifying conceptual relations with a given model
Identifying phenomenon (phen idfn)	To identify a phenomenon: identifying its defining features, and the problem associated
Identifying the model	Identifying the model that would be useful in explaining the

(model idfn)	phenomenon
Identifying the variables (var idfn)	Identifying the relevant variables that are useful for explaining a phenomenon
Identifying variable relations (var rel)	Identifying the relationships between identified variable in a model
Measuring the variable (var mes)	To know how the variable can be measured in an experimental setup
Regulating variables (var regul)	To know how to regulate the variables in an experimental setup; to satisfy ceteris paribus conditions
Designing Experiments (design)	Designing experiments that would produce data to either A) construct observational generalisations, or B) validate predictions of models
Identifying counter-examples (count eg)	To identify counter examples to a model
Procedural knowledge (rout proceed)	To apply routine procedure templates to solve a problem
Technical skills (tech skills)	The technical know-how's to carry out an experiment

Data Modelling (data mod)	Modelling collected data, with the purpose of description or construction of observation generalisations
Defining (def)	The ability to define concepts/variables/relations

Description of primitives

Model : A system of coherent and logically consistent propositions that seek to explain a phenomenon

Phenomenon : any event under consideration for the purpose of description or explanation

System of

Propositions : A set of propositions that are coherent (that they relate to the same object)

Premise : (of a model) the definitions, relationships that a model assumes to be true, from which inferences (predictions) are derived

: (of a problem) equivalent to boundary conditions imposed on the model in the context of a problem

Problem : a situation or task that has at least one unknown variable or unknown relationship between variables

Concept : A category of ideas that are idealised and abstracted

Variable : A function of concepts in a model, establishes correspondence between concept and observation

Appendix 2

The edges in the graph are representations of the following sentences. They have the form: *<ability> has the pre-requisite <ability>*. The definitions of the abilities are listed in Appendix 1.

The ability to evaluate a proposition has the pre-requisite ability to check for logical consistency
The ability to evaluate a proposition has the pre-requisite ability to compare models

The ability to construct models has the pre-requisite ability to postulate
The ability to construct models has the pre-requisite ability to reason

The ability to solve problems has the pre-requisite ability to identify the phenomenon
The ability to solve problems has the pre-requisite ability to identify the model
The ability to solve problems has the pre-requisite ability to apply routine procedures

The ability to check logical consistency has the pre-requisite ability to reason
The ability to reason has the pre-requisite ability to identify the logical relationship between propositions
The ability to reason has the pre-requisite ability to identify the premises
The ability to reason has the pre-requisite ability to identify the model

The ability to identify model has the pre-requisite ability to define
The ability to identify model has the pre-requisite ability to identify instances
The ability to identify model has the pre-requisite ability to identify conceptual relations
The ability to identify model has the pre-requisite ability to represent the conceptual relations
The ability to identify model has the pre-requisite ability to identify the variables
The ability to identify model has the pre-requisite ability to identify variable relations

The ability to compare models has the pre-requisite ability to identify counter examples
The ability to compare models has the pre-requisite ability to identify the assumptions

The ability to experiment has the pre-requisite ability to design experiments
The ability to experiment has the pre-requisite ability to measure the variables
The ability to experiment has the pre-requisite ability to regulate the variables

The ability to design experiments has the pre-requisite ability to identify variable relations

The ability to postulate has the pre-requisite ability to define
The ability to postulate has the pre-requisite ability to model data

The ability to model data has the pre-requisite ability to define
The ability to model data has the pre-requisite ability to collect data
The ability to model data has the pre-requisite ability to identify variable relations

The ability to identify variable relations has the pre-requisite ability to define
The ability to identify variable relations has the pre-requisite ability to identify conceptual relations

The ability to measure variables has the pre-requisite ability to identify variable relations
The ability to measure variables has the pre-requisite of technical skills

The ability to regulate variables has the pre-requisite ability to identify variables
The ability to regulate variables has the pre-requisite of technical skills

References

Asamoah, D., & Ocansey, M. K. K. (2019). *Item Discrimination and Distractor Analysis: A Technical Report on Thirty Multiple Choice Core Mathematics Achievement Test Items*.

Bertucio, B. (2017). The Cartesian Heritage of Bloom's Taxonomy. *Studies in Philosophy and Education*, *36*(4), 477–497. https://doi.org/10.1007/s11217-017-9575-2

Bezuidenhout, M. J., & Alt, H. (2011). "Assessment drives learning": Do assessments promote high-level cognitive processing? *South African Journal of Higher Education*, *25*(6), 1062–1076.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems*, *Special Issue on Linked Data*.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. 1: Cognitive domain* (26. print). Longman Inc.

Boumans, M. (1999). Built-in justification. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social sciences*. Cambridge University Press.

Chaudhuri, N. B., Dhar, D., & Yammiyavar, P. G. (2021). Do Design Entrance Exams in India Really Test Creative Aptitude? An Analytical Study of Design Tests Conforming Creativity Benchmarks. In A. Chakrabarti, R. Poovaiah, P. Bokil, & V. Kant (Eds.), *Design for Tomorrow—Volume 2* (Vol. 222). Springer Singapore. https://doi.org/10.1007/978-981-16-0119-4

Chowdhury, A. R., N. K. S., Prakash, V., & Das, S. (2023). *A comparative analysis of the cognitive levels of Science and Mathematics secondary school board examination questions in India*. https://doi.org/10.5281/ZENODO.8115778

Constantinou, C. P., Nicolaou, C. Th., & Papaevripidou, M. (2019). A Framework for Modeling-Based Learning, Teaching, and Assessment. In A. Upmeier Zu Belzen, D. Krüger, & J. Van Driel (Eds.), *Towards a Competence-Based View on Models and Modeling in Science Education* (Vol. 12). Springer International Publishing. https://doi.org/10.1007/978-3-030-30255-9

Deshpande, S. (2006). Exclusive Inequalities: Merit, Caste and Discrimination in Indian Higher Education Today. *Economic and Political Weekly*, *41*(24), 2438–2444.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262–271. https://doi.org/10.1016/j.shpsa.2010.11.034

Madan, A. (2007). Sociologising Merit. *Economic and Political Weekly*, *29*, 3044–3050.

Mohanan, K. P. (2010). Entrance examinations for science and technology. *Current Science*, *99*(10), 1321–1323. JSTOR.

Morgan, M. S., & Morrison, M. (Eds.). (1999). *Models as mediators: Perspectives on natural and social sciences*. Cambridge University Press.

Nagarjuna, G., Kharatmal, M., & Nair, R. (2010). Building a Dependency Network for Teaching-Learning of Conceptual Structures. *Proceedings of the International Workshop CS-LTA 2010*.

Nersessian, N. J. (1999a). Model-Based Reasoning in Conceptual Change. In L. Magnani, P. Thagard, & N. J. Nersessian (Eds.), *Model-Based Reasoning in Scientific Discovery* (1st ed.). Springer Science+Business Media. https://link.springer.com/book/10.1007/978-1-4615-4813-3

Nersessian, N. J. (1999b). *Model-Based Reasoning in Scientific Discovery* (L. Magnani, P.

Thagard, & N. J. Nersessian, Eds.; 1st ed.). Springer Science+Business Media. https://link.springer.com/book/10.1007/978-1-4615-4813-3

Patwardhan, B., Mohanan, K. P., Mohanan, T., Ramakrishnan, V., Grewal, R. S., Krishna, R., Patel, V. S., Sancheti, S., Ghosh, R., Mittal, P., Kawday, D., Sharma, S. C., Sridhar, M. K., Kamat, V., Prasad, V. S., Mehta, S. R., Jauhri, A., Vaidhyasubramaniam, S., Deshpande, M., … Sahasrabudhe, V. (2022). *Re-imagining Assessment and Accreditation in Higher Education in India*. National Assessment and Accreditation Council.

Punjabi, S. (2020). Is Shadow Education Becoming the 'New' Formal? Effects of Pedagogical Approaches of IIT-JEE Coaching on School Education in the City of Delhi. *Contemporary Education Dialogue*, *17*(1), 14–44. https://doi.org/10.1177/0973184919885485

Rost, M., & Knuuttila, T. (2022). Models as Epistemic Artifacts for Scientific Reasoning in Science Education Research. *Education Sciences*, *12*(4), 276. https://doi.org/10.3390/educsci12040276

Russo, M. (2011). Aptitude testing over the years. *Interpreting. International Journal of Research and Practice in Interpreting*, *13*(1), 5–30. https://doi.org/10.1075/intp.13.1.02rus

Sawyer, R. K. (2006). *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Soozandehfar, S. M. A., & Adeli, M. R. (2016). A Critical Appraisal of Bloom's Taxonomy. *American Research Journal of English and Literature*, *2*, 1–9. https://doi.org/10.21694/2378-9026.16014

Turnbull, J. M. (1989). What Is… Normative versus Criterion-referenced Assessment. *Medical*

Teacher, *11*(2), 145–150. https://doi.org/10.3109/01421598909146317

Wong, A. S. T., & Park, C. B. (2010). A Case Study of a Systematic Iterative Design Methodology and its Application in Engineering Education. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.v0i0.3148