Assessing concept possession as an explicit and social practice

Alessia Marabini
University of Bologna
alessiamarabini7@gmail.com

Luca Moretti
University of Aberdeen/MCMP
l.moretti@abdn.ac.uk

Final Draft

ABSTRACT

We focus on issues of learning assessment from the point of view of an investigation of philosophical elements in teaching. We contend that assessment of concept possession at school based on ordinary multiple-choice tests might be ineffective because it overlooks aspects of human rationality illuminated by Robert Brandom's inferentialism—the view that conceptual content largely coincides with the inferential role of linguistic expressions used in public discourse. More particularly, we argue that multiple-choice tests at schools might fail to accurately assess the possession of a concept or the lack of it, for they only check the written outputs of the pupils who take them, without detecting the inferences actually endorsed or used by them. We suggest that school tests would acquire reliability if they enabled pupils to make the reasons of their answers or the inferences they use explicit, so as to contribute to what Brandom calls the game of giving and asking for reasons. We explore the possibility to put this suggestion into practice by deploying two-tier multiple-choice tests.

KEYWORDS: school assessment; multiple-choice tests, two-tier multiple-choice tests, concept mastery, concept possession, inferentialism, material inference, non-monotonic inference, Robert Brandom

1. Introduction[1]

In this paper we focus on an issue of learning appraisal from the point of view of an investigation of philosophical elements in teaching. We submit that assessment of concept possession at school based on ordinary multiple-choice (MC) tests might be ineffective because it disregards aspects of human rationality illuminated by Robert Brandom (1994 and 2000)'s *inferentialism*—the view that conceptual content largely coincides with the *inferential role* of linguistic items used in public discourse.

---

[1] We are grateful to Jan Derry, Fed Luzzi, Paul Standish, Shone Surendran, the audience of a research seminar of the Institute of Education of UCL, and two anonymous reviewers of this journal for comments on a draft of this paper.

Drawing from Sellars (1953) and Brandom (2000), we suggest that conceptual content, conceived of as an inferential role, goes hand in hand with the ability to draw certain inferences. We contend that MC-tests might fail to accurately assess the possession of a concept or the lack of it, for they only check the written outputs of the pupils, without detecting the inferences actually endorsed or used by them. We propose that school tests would become more reliable if they enabled the pupils who take them to make the reasons of their answers or the inferences they use explicit, so as to contribute to the social practice that Brandom call the *game of giving and asking for reasons*. We explore the possibility to put this proposal into practice by deploying *two-tier* multiple-choice tests.

The relevance of Brandom's inferentialism to philosophy of education has been investigated and illuminated by Derry (2013 and 2017). We agree with Derry's thesis that inferentialism plays— or should play—a pivotal role in this discipline. Our paper is an attempt to describe an application of inferentialism unexplored in Derry's work, which supplies fresh evidence for that thesis.

The paper is organized as follows: §2 introduces the problem of assessing pupils' concept mastery by presenting two paradigmatic cases. § 3 outlines Brandom's view of inferentialism and concept possession. § 4 elucidates the notion of materially valid inference. § 5 delivers our inferentialist diagnosis of the problematic cases of assessment of concept mastery. § 6 suggests a remedy. § 7 concludes the paper.

2. A problem of assessing pupils' concept mastery

To introduce the topic we detail two paradigmatic cases of test assessment based on MC-questions that give rise to the difficulties we are interested in. These cases are based on two scientific literacy test items to assess the learning outcomes of pupils. The first test item is part of an annual test for elementary schools prepared by the *Office of State Assessment* of the *New York State Education Department* and it is meant to be taken by 4th grade pupils—so, nine/ten-year old children. The

second test item is part of an annual test for primary and secondary schools, called INVALSI, prepared by the Italian *National Institute for the Education System Evaluation*. This test item is meant to be taken by pupils at level two—so, seven-year old children. Since the problems we describe are very basic and general in character, they are likely to affect similar school tests in other countries.

Ordinary MC-test items present students with a stimulus for a response—called *stem*—typically including a description and a question, and a few suggested answers, one of which—the *key*—is the correct one. The other, incorrect answers are called *distractors*.[2]

This is our first test item.

(SHADOW)
There is a shadow under a tree. Which form of energy must be present for the shadow to occur?
A. Heat.   *B*. Light.   *C*. Sound.   *D*. Mechanical.[3]

According to the answer key, the competent pupil who masters the concept of light must select *B*. We can wonder, however, whether a pupil's answering *B* is inevitably evidence that the pupil masters the concept of light. The answer is negative. Suppose Claire answers *B*. We are invited to think that this is evidence that Claire masters the notion of light. The assumption is that Claire's selecting *B* attests that she knows that the form of energy that must be present for a shadow to occur is just light. Despite this, it is easy to realize that Claire's answer might result from her *not* possessing this knowledge. For instance, suppose Claire's answering *B* depends on her believing that the forms of energy that must be present for a shadow to occur are *both* light *and* heat, except for when the shadow is cast by a tree—like in SHADOW's scenario—in which case the presence of light *alone* suffices for the shadow to occur. False beliefs of this sort are not infrequent in children

---

[2] For other ordinary formats see Haladyna (2004: Ch. 4).
[3] Cf. *2013 Grade 4 Elementary Level Science Test*, p. 4. Downloadable at:
http://www.nysedregents.org/Grade4/Science/home.html

who gradually learn concepts.[4] It is intuitive that Claire's answering *B* in this case would *not* be

evidence that she possesses the concept of light.

This is the second test item.

> (BEACH)
> Carl goes to the beach with a paper boat, a small stone and a balloon. Which one of
> these objects will sink in the water?
>
> *A*. The paper boat.   *B*. The small stone.   *C*. The balloon.[5]

The competent pupil who masters the concept of sinking (as opposed to floating) must select *B*. So

if a child selected *C* (or the other distractor) we would be invited to conclude that the pupil doesn't

master this notion. Yet this conclusion might prove erroneous in some cases. Suppose Claire

answers *C*. Also suppose that Claire's answer is dictated by this: she imagines that once Carl arrives

at the beach, he will put his small stone on the paper boat, which is capable of sustaining its weight

without sinking. Furthermore, Claire imagines that Carl loves to deflate balloons, fill them with

sand and throw them into the sea to watch them sink, for Claire likes to do the same. It is not

implausible that a seven-year old could imagine this scenario in her first experience with a test of

this type. One might blame Claire for her imagination.[6] Yet BEACH doesn't require pupils not to

be imaginative. It is intuitive that Claire's answering *C* is *not* evidence that she doesn't master the

concept of sinking.

Problems like those affecting SHADOW and BEACH may affect other MC-test items. We

can say that the difficulties emerge in two sets of circumstances: first, when the pupil who takes a

test item entertains false beliefs that would seem to preclude her from mastering the required

concept but the false beliefs are undetected because the pupil ends up selecting the test item's key;

---

[4] Many examples of this type can be found in the web site *I used to believe* (https://www.iusedtobelieve.com). This is a
collection of bizarre ideas that adults had when they were children. The section "Science" is of particular interest.
[5] Cf. *INVALSI*, *Anno Scolastico 2006–2007*, *Scienze classe seconda*, p. 9. Downloadable at:
http://www.icsanstino.gov.it/progetti-e-iniziative/prove-invalsi/14-progetti-e-iniziative/286-per-allenarsi-scienze-
scuola-primaria-classe-seconda.
[6] We mean *rational* imagination—e.g., in the sense elucidated by Byrne (2005)—rather than unconstrained fantasy.

second, when the pupil does master or may master the required concept but ends up selecting a distractor because she fills in details of the scenario sketched in the test item's stem in an unexpected way.

To make this diagnosis clearer-cut, we need to clarify what it means to master a concept when this mastery has essentially to do with the ability to assert the correct conclusion. By 'correct conclusion' we mean the one based on the content of the concepts involved in the question asked in a test—this is the conclusion that pupils are required to infer according to the test's answer key. This issue takes us to a related one: clarifying our ability to draw inferences. Brandom's inferentialism and Brandom and Sellars' conception of a materially correct inference are particularly useful for these purposes.

3. Brandom's inferentialism

Broadly speaking, inferentialism is the view according to which having a conceptual content is, for a linguistic expression, having an *inferential role*. A (quite trite) analogy between language and chess is helpful to give a preliminarily illustration of this view: just as the fact that a piece of wood is a queen or a rook is a matter of the moves it is allowed to make on the chessboard, the fact that a type of ink spot or sound has a specific content is a matter of the inferences in which it is permitted to feature. Inferentialism is opposed to semantic *representationalism*—a more traditional view according to which linguistic expressions have meaning essentially because they represent things. Advocates of this view or family of views are for instance Devitt (1981), Etchemendy (1990) and Fodor (1998).

The term 'inferentialism' was introduced in philosophy by Robert Brandom to refer to his own conception of language and meaning.[7] As a pragmatist, Brandom thinks of language as a tool to play various *activities*, or *language games*—to use an expression introduced by Wittgenstein (see

---

[7] Other philosophers—e.g. Sellars (1949, 1953, 1974) and Boghossian (1996)—have embraced similar conceptions, though less developed or applied only to circumscribed areas.

Brandom, 1994 and 2000). The most central of these language games is—according to him—the one of *giving and asking for reasons*. This activity is especially important because it unveils our rational nature of discursive and concept-using creatures (cf. 2000: 81). Brandom intends to provide a systematic account of how the use of linguistic expressions within the game of giving and asking for reasons can confer conceptual content on those expressions (cf. 2000: 1-15).

Brandom's key thesis is that our language enables us to play this important game because it is *inferentially* articulated. (We analyse Brandom's notion of inference in the next section.) In fact, note that to be able to give reasons for a claim, one must be able to make other claims that work as *premises* of *arguments*. More specifically, Brandom maintains that our language is inferentially articulated because asserting that *P* necessarily comes with certain inferential *commitments* and *entitlements*. A subject *S* asserting that *P commits S* to asserting the claims *S* can infer from *P*. For instance, *S*'s asserting 'This is a fish' commits *S* to asserting 'This lives in water' and 'This is not a mammal'. Furthermore, *S*'s asserting that *P entitles S* to the mentioned commitments and makes *S* accountable for *S*'s assertion that *P* and the ensuing entitlements. This accountability commits *S* to exhibiting *P* as the conclusion of some inference from other claims (cf. 2000: 11 and 43-44). For instance, *S*'s asserting 'This is a fish' commits *S* to inferring this claim from other claims, such as 'This is a salmon'.

Brandom holds that this web of inferential commitments and entitlements associated with the (potential) assertion of a given expression defines the inferential role of the expression and makes it have a specific content. Inferentialism is, first of all, a *propositional* doctrine because inferential roles primarily determine the *propositional contents* of claims. The contents of their constitutive parts—e.g. singular terms and predicates—are determined only in consequence. For our circumscribed purposes, it is sufficient to say that the meaning of a noun or predicate—such as 'house' or 'rabbit'—is determined by the inferential roles of the sentences in which these

expressions occur. If these inferential roles changed, the meaning of these sub-sentential expressions would change too (cf. 2000: 13, 30 and § 4).

For Brandom, a subject *S* possesses the conceptual content of a given statement only if *S* endorses and thus would carry out—in the appropriate circumstances—the inferences required by the associated commitments and entitlements. This requires *S* to be able to make many other assertions associated with further commitments and entitlements. Brandom's inferentialism is thus a form of meaning *holism*, for the inferential commitments and entitlements of the speaker are such that the speaker can master one concept only if she masters many other inferentially correlated concepts (cf. 2000: 15-16 and 29).

Brandom's inferentialism can be described as a form of *social* and *normative* linguistic pragmatism because it is the web of inferential commitments and entitlements associated with the assertion of a sentence that grounds the sentence's inferential role. A speaker can play the game of giving and asking for reasons only if her own linguistic commitments and entitlements are acknowledged by her interlocutors and she acknowledges those of her interlocutors within her linguistic community. Furthermore, commitments and entitlements like those described are—for Brandom—*normative statuses* that people continuously acquire while communicating with one another within a linguistic community. In general, our being members of a community requires us to abide by a rich variety of social norms that depend on the normative statuses of the community's members. In accordance with this, competently speaking a language requires us to obey norms depending on our *linguistic* commitments and entitlements (cf. 2000: 79-84).

Take now a claim—such as 'This cube is red'—*directly elicited by a perception*. One might suppose that the conceptual content of basic claims of this sort is fully determined by their ability to represent something—e.g. the fact that this cube is red—independently of their inferential roles. But Brandom's *strong* form of inferentialism denies this (cf. 2000: 28 and 219-220). Brandom's inferentialism concedes that the conceptual content of these basic claims is *partly* determined by

their circumstances of appropriate application. However, Brandom maintains that these claims too, in order to have conceptual content, must be *inferentially articulated* (cf. 2000: 47-48 and 108-109). The capacity to inferentially articulate non-inferentially elicited claims is what marks the difference between ordinary humans and "speaking" animals—such as parrots—and simple tools—like thermostats and photocells. Consider a parrot saying 'This cube is red' before a red cube, or a thermostat indicating the correct temperature. Whereas the answers by the parrot and the thermostat provide a mere *responsive* classification of the data, the answers by humans, in similar circumstances, would offer a *conceptual* classification of them. Thermostats and parrots—unlike humans—don't *understand* their own responses. This happens because only humans, but not parrots and thermostats, could *infer* from their non-inferentially produced claims many other claims. For example, from 'This cube is red', ordinary subjects could infer 'There is something red' or 'This cube isn't blue' (cf. 2000: 47-48).

Let's dwell on an important notion used by Brandom to characterize his view. According to Brandom, *making something explicit* is the same as *conceptualizing* it—namely, putting it into a statement and so into something that both *serves as* and *stands in need of* a reason (cf. 2000: 11 and 16). Brandom individuates various implicit things that a speaker *S* can make explicit by playing the game of giving and asking for reasons. The first thing is what is made explicit by *S*'s statements— namely, *propositions* or *possible facts*. Two further things are what remains implicit in *S*'s explicit statements—namely, their *consequences* and the *reasons* for them. *S* can make these consequences and these reasons explicit by asserting them. A fourth thing that can be made explicit by *S* is the *inferences* that *S* implicitly commits herself to in making a claim. Suppose that an inference *S* implicitly commits herself to in claiming that *P* is '*P*. Therefore, *Q*'. If *S*'s logical language is rich enough to contain conditionals, *S* can make this inference explicit by stating 'If *P* then *Q*' (cf. 2000: 11, 14, 18-20).

4. Material inference

For Brandom, what constitutes the meaning of a linguistic expression is essentially its inferential role. To further clarify this thesis we need to explain what Brandom means by 'inference'. Brandom holds that conceptual contents are to be understood in terms of both the *formally* valid and the *materially* valid inferences they underwrite. Hereafter, we will use 'formal inference' and 'material inference' as abbreviations of respectively 'formally *valid* inference' and 'materially *valid* inference'.

Sellars (1953) introduced the notion of material inference by contrasting it with the one of formal inference. A formal inference is a good inference in virtue of its *logical form*. A material inference is a good inference in virtue of the *non-logical* contents of its premises and conclusion. A formal inference is, for example, any instance of the inference schema *Modus Ponens*:

> If P, then Q.
> P.
> Therefore, Q.

Take P = 'Socrates is a man' and Q = 'Socrates is mortal'. The following inference is formally valid:

> (SOCRATES)
> If Socrates is a man, then Socrates is mortal.
> Socrates is a man.
> Therefore, Socrates is mortal.

Indeed, (SOCRATES) appears correct independently of the meanings of P and Q. Its correctness rests on its being an instance of *Modus Ponens*.

Material inferences are for example the following three:

> (LIGHTNING)
> Lightning is seen now.
> Therefore, thunder will be heard soon.

(RED BALL)
The ball is red all over.
Therefore, the ball is not yellow.

(MATCH)
I strike this dry, well-made match.
Therefore, it will light.

When taken at face value, (LIGHTNING), (RED BALL) and (MATCH) don't appear to possess

any logical form capable of grounding their validity.[8] These three inferences are nevertheless good

and they are normally acknowledged to be so. Sellars would say that these arguments are *materially*

good, in the sense that they are valid in virtue of the specific *contents* of their premises and

conclusions. Sellars would also say that one's endorsing these inferences—regardless of one's

having any *logical* competence—is required by one's grasping or mastering those contents.

Brandom (2000) follows Sellars in considering inferences like the above ones to be

materially valid. He emphasizes that the same arguments are sometimes interpreted as *disguised*

formal inferences. On this interpretation, inference like the above ones are taken to be *enthymemes*–

–i.e. arguments in which a premise is not explicitly stated. For example, (LIGHTNING) is

understood as implicitly involving the conditional premise 'If lightning is seen, then thunder will be

heard soon'. When this "suppressed" premise is explicitly stated, the inference becomes an instance

of *Modus Ponens* and, as such, it proves *formally* valid. Brandom calls this *the formalist approach*

to inference (cf. 2000: 53).

Brandom rejects the formalist approach essentially because construing material inferences

as enthymematic formal inferences proves less straightforward than it might appear at first.[9]

Formal inferences are in fact *monotonic*. A valid inference is monotonic just in case if we add a

premise to the ones already in place, the inference remains valid (cf. 2000: 87). For instance, if we

add the statement 'Today is Saturday' or 'Socrates is a Greek' to the premises of (SOCRATES), the

---

[8] Although (RED BALL) looks *a priori* valid, its validity doesn't rest on its *logical form*.
[9] For further arguments see Brandom (2000: 53, 55 and 85-86).

resulting inference is valid. Brandom stresses that—unlike formal inferences—*material*

inferences are often *non-monotonic* (cf. 2000: 88). He offers this example:

> (1) I strike this dry, well-made match. Therefore, it will light. (*P*. Therefore, *Q*)
>
> (2) *P* and the match is in a very strong electromagnetic field. Therefore, it will not light. (*P & R*. Therefore, *Not-Q*)
>
> (3) *P* and *R* and the match is in a Faraday cage. Therefore, it will light. (*P & R & S*. Therefore, *Q*)
>
> (4) *P* and *R* and *S* and the room is evacuated of oxygen. Therefore, it will not light. (*P & R & S & T*. Therefore, *Not-Q*)

Arguments (1)-(4) show that it is possible to produce hierarchies of material inferences *with oscillating conclusions* by merely *adding further premises*. (1)-(4) exemplify the claim that material inferences in ordinary discourse and science are often non-monotonic. Given this, many material inferences cannot be construed as formal inferences in disguise. For instance, if (1) were an enthymematic formal inference, it should be monotonic. But the apparent goodness of (2) indicates that (1) is not so (cf. 2000: 88).

Brandom concedes that those who adopt a formalist approach might respond by invoking *ceteris paribus* clauses. Take again (1). An advocate of this approach—call her *the formalist*— might claim that the premise implicit in (1) says precisely this: '*Ceteris paribus*, if I strike this dry, well-made match, it will light'. '*Ceteris paribus*' means 'In the absence of disturbing factors'. The formalist would also claim that the only premise *explicit* in (1) is to be read: '*Ceteris paribus*, I strike this dry, well-made match'. So interpreted, (1) appears formally valid. The formalist will conclude that since the match being in a very strong electromagnetic field counts as a disturbing factor, the premise of (2) describes a case in which the ceteris paribus clause implicit in the premise of (1) remains unsatisfied. So (2) doesn't show that (1) is non-monotonic.

A notorious problem of ceteris paribus clauses is, however, that they cannot be replaced by explicit descriptions of disturbing factors, for there is no way to specify these factors in advance. A

consequence is that a conditional such as 'Ceteris paribus, if I strike this dry, well-made match, it will light' is ultimately vacuous and trivially true. For it means: 'If I strike this dry, well-made match, it will light *unless it won't do it for some reason*'.[10] Brandom's conclusion is this: claiming that our implicit acceptance of empty conditionals of this type explains why *monotonic* formal inferences can instantiate hierarchies like (1)-(4) is more ad hoc and thus less plausible than admitting that the same inferences are genuinely *non-monotonic* and irreducibly material (cf. 2000: 88).

We find Brandom's inferentialism plausible. It appears to us to be intuitively true that the conceptual content of our claims is tightly linked to their inferential roles. Furthermore, it seems to us that many of our inferences just are non-monotonic material inferences. For reason of space, we cannot provide a defence of inferentialism here. For arguments in support of it and responses to criticism, see Jorgensen (2008), Peregrin (2009 and 2012), and Murzi and Steinberger (2017). The reader of this paper is free to interpret our central claims in a *conditional* fashion: *if* Brandom's inferentialism is correct, *then* we acquire a clear-cut understanding of why the assessment of tests at school may mislead teachers, and we have a suggestion for a remedy at hand.

5. Assessment of tests at school: an inferentialist diagnosis

In § 2 we have made two intuitive claims: first, the fact that a pupil *passes* the test SHADOW may not be evidence that she masters the notions of light; second, the fact that a pupil *fails* the test BEACH may not be evidence that she doesn't master the notion of sinking. Brandom's inferentialism helps us clarify and substantiate both intuitions.

Consider again SHADOW and Claire's selecting the test item's key *B*. Once inferentialism is in place, we can see with clarity why, despite her "correct" answer, Claire doesn't master the concept of light. This is so because she doesn't endorse certain inferences that shape the inferential

---

[10] Cf. Lange (1993) and Hempel (1988).

role of the sentences in which the expression 'light' occurs. The problem for Claire is that these inferences also determine—indirectly—the meaning or concept of 'light'. For example, given her misconception of light,[11] Claire cannot endorse this *correct* material inference that partly defines the meaning of 'light':

> There is a shadow.
> Therefore, the form of energy that must be present is light.

The *incorrect* material inference that Claire endorses in its place is this:

> There is a shadow.
> Therefore, the forms of energy that must be present are light and heat.

Why does Claire nevertheless select answer *B*? Because once she understands the features of the scenario sketched in SHADOW's stem, Claire runs the following inference:

> There is a shadow.
> It is cast by a tree.
> Therefore, the form of energy that must be present is light.

There is no conflict between this inference and the former one, which has an incompatible conclusion. Material inferences like these are in fact non-monotonic.

Consider now BEACH and Claire's selecting the distractor *C*. Once inferentialism is in place, we can unmistakably see why, despite her "incorrect" answer, Claire may still master the concept of sinking. This is so because Claire's answer is compatible with her endorsing all inferences that determine the inferential roles of our sentences embedding 'sink' and its derivatives. These inferences determine, indirectly, the concept or meaning associated with this sub-sentential expression. For instance, Claire may still coherently endorse the following material inferences presupposed by the answer key of BEACH:

---

[11] That it can produce shadows without the help of heat only in presence of trees.

This is a paper boat.
Therefore, it won't sink in the water.

This is a small stone.
Therefore, it will sink in the water.

This is a balloon.
Therefore, it won't sink in the water.

Why does Claire nevertheless select answer *C*? Because once she has come up with a non-implausible scenario involving a paper boat, a little stone and a balloon, coherent with all information supplied in BEACH, she runs the following three correct material inferences:

This is a paper boat.
Carl has put a small stone on it.
Therefore, it won't sink in the water.

This is a small stone.
Carl has put it on a paper boat.
Therefore, it won't sink in the water.

This is a balloon.
Carl has filled it with sand.
Therefore, it will sink in the water.

These inferences are all good and their premises are true in the envisaged scenario. There is no conflict between the last two inferences and the last two of the triplet of considered before. For material inferences of this type are non-monotonic.

Brandom's view enables us to provide a simple and unified diagnosis of the problems afflicting both SHADOW and BEACH. The difficulty afflicting these tests is that their answer keys don't enable us to check the inferences actually used or endorsed by the pupils who take the test items, though this information would be relevant to correctly assess whether or not these pupils master the relevant concepts. For instance, the answer key of SHADOW doesn't enable us to detect the following *incorrect* inference endorsed by Claire: 'There is a shadow. Therefore, the forms of energy that must be present are light and heat'. By the same token, the answer key of BEACH

doesn't enable us to detect any of the *correct* inferences endorsed or used by Claire that shape the inferential roles of the sentences involving the expression 'sink'. The only things that the answer keys of SHADOW and BEACH enable us to check are Claire's *written outputs* (e.g. her chosen answer to the test questions), which turn out to be misleading in the appraisal of her concept possession.

One might propose to settle these difficulties by simply *sharpening* the questions in the stems of these test items in the attempt to contain possible misguiding effects of beliefs of the pupils who take the tests and their invasive imagination. For instance, one might re-phrase (SHADOW) into this more succinct question:

> (*SHADOW*)
> Consider a shadow: which form of energy must be present for it to occur?

We agree that the implementation of this proposal would help us reduce the number of problematic cases. But we think that it wouldn't resolve all problems. For children aren't always capable of recognizing what is actually presupposed in a question, quite independently of its sharpness—this is something that pupils learn in time. Furthermore, children who undergo the multifaceted process of concept acquisition are always liable to developing unexpected and odd beliefs. For instance, it wouldn't be particularly surprising if Claire answered 'Light' to the question in *SHADOW* because she entertained the false belief that light doesn't require heat to produce a shadow *when it is surrounded only by air*, but it needs heat to do so when it is surrounded by other elements—for example, glass or water. And it wouldn't be surprising if Claire thought that a presupposition of *SHADOW* is that what casts the shadow is in the open air.

We believe that the difficulties that affect SHADOW and BEACH somewhat afflict all ordinary MC-test items. For all test items of this type check only the written outputs of pupils but not their inferences. The problem is that a pupil's written output is something essentially similar to

a parrot's disposition to say 'That's red' or 'That's a banana' in given circumstances. On the other hand, a concept is an *inferentially articulated* entity irreducible to a mere disposition to make this or that specific claim. Testing whether one possesses a given concept requires checking whether one actually draws and would be ready to draw inferences of a given type. This can hardly be done by simply checking written responses like those of SHADOW and BEACH.

6. A proposal: deploying two-tier MC-tests

What solution or remedy should we recommend in light of our last observations? To begin with, we think that it is unlikely that the outcomes of MC-tests *alone* could provide an adequate representation of pupils' concept possession. Information suitable to appraise pupils' concept possession should be sought from a variety of different sources, some of which capable of offering a more direct and detailed picture of the inferences that pupils draw and would be ready to draw. We refer—for example—to free response tests (in form of essays or shorter answers), oral examinations, and Socratic dialogue and community of inquiry.[12] We think that the use of MC-tests for the purpose to assess pupils' concept possession shouldn't nevertheless be discontinued. For MC-tests have practical advantages—for example, they are less time-consuming for both the assessed and the assessor than tests of other types. Furthermore, they permit coverage of many topics at once.[13] Our view is that *certain types* of MC-tests are more suitable than others to appraise concept possession.

Consider SHADOW and BEACH once again. If Claire made it explicit that the reason why she selected *B* in SHADOW is that the shadow is cast by a tree, it would become apparent that she doesn't master the notion of light. For if Claire actually endorsed the inferences that shape the notion of light, she should acknowledge that whether or not the shadow is cast by a tree is utterly

---

[12] See Dunlop et al. (2015) for a description of how the method of Socratic dialog and community of inquiry can be applied to primary school science.

[13] For a comprehensive list of acknowledged advantages and disadvantages of MC-tests see Simkin and Kuechler (2004).

irrelevant to the form of energy that must be present for the shadow to occur. Similarly, if Claire described the inferences drawn to select *C* in BEACH, it would be plain that her answer doesn't indicate that she doesn't master the notion of sinking and floating. For those inferences are compatible with those that define the notion of sinking. The general conclusion we draw is this: MC-tests would become more reliable in appraising concept mastery if the pupils taking them were asked to make the reasons of their answers or the inferences that they draw explicit, so as to contribute to what Brandom's calls the game of giving and asking for reasons. One possible way to put this suggestion into practice is deploying *two-tier* MC-tests.

Students' scientific *misconceptions* have been an enduring interest in philosophy and science of education (cf. Allen 2010). Research data show that pupils typically go to science classes with pre-instructional concepts and beliefs about the topics to be taught, and that this often interferes with their learning and performance on tests (cf. Duit and Treagust 2003). A criticism of *ordinary* MC-tests—such as SHADOW and BEACH—is that they don't afford sufficiently deep insight into pupils' conceptions, as they are unable to discriminate between correct responses and *fortuitously* "correct" responses—namely, cases in which the test's key is selected *for a wrong reason* due to an interfering misconception (cf. Rollnick and Mahooana 1999). Claire's selecting *B* in SHADOW exemplifies this possibility. Claire can be said to entertain a misconception of light that leads her to incidentally select the "correct" answer *B*. Tamir (1971, 1989 and 1990) proposed to use enhanced MC-tests that required students to *justify* their answers by supplying reasons for them. He found that these tests were able to detect fortuitously "correct" responses; this happened when the reason given for the "correct" answer was recognized to be a mistaken one. The effectiveness of MC-tests of this type as diagnostic tools has been corroborated by subsequent investigation (cf. Tsai and Chou 2002 and Wang 2004). This has elicited the development of two-tier MC-tests.

The first tier of any two-tier MC-test item is nothing but a standard MC-test item, such as SHADOW or BEACH. It consists of a stem enclosing a question followed by a list of possible answers whose only one is correct. The second tier consists of a new question asking for a justification of the first-tier answer. This question is followed by a list of possible reasons, explanations or arguments, depending on the test. Only one of these responses is correct. The distractors may be based on misinterpretations that are known to be widespread among the test takers. Some two-tier MC-tests allow students to write down their own second-tier answer if it is not included in the list. A student passes a two-tier MC-test item when she responds to both its questions correctly (cf. Haladyna 2004: 143-145 and Gurel et al. 2015).[14]

The following MC-test items are possible two-tier variants of respectively SHADOW and BEACH:[15]

(SHADOW*)
(1) There is a shadow under a tree. Which form of energy must be present for the shadow to occur?
A. Heat.   *B*. Light.   *C*. Sound.   *D*. Mechanical.

(2) Why?
E. Because the shadow would shelter us from the heat.
F. Because the shadow is caused by the tree's obstructing a light source.
G. Because the shadow is extracted from the tree.
[*The correct answer may not be in this list. Your can write your own answer below*.]
H. Because…

(BEACH*)
(1) Carl goes to the beach with a paper boat, a small stone and a balloon. Which one of these objects will sink in the water?
A. The paper boat.   *B*. The small stone.   *C*. The balloon.

(2) What is your reasoning?
D. It has sharp edges. Therefore, it will sink in the water.
E. It is heavy. Therefore, it will sink in the water.
F. It is not flat. Therefore, it will not float on the water.

---

[14] Some authors have proposed replacing two-tier MC-tests with *three*- or *four*-tier MC-tests to further increase their reliability as diagnostic tools. The students sitting a three- or four-tier MC-test item are asked to specify the strength of their confidence in, respectively, their second-tier answer or both their first- and second-tier answers (cf. Gurel et al. 2015: 995-999). We don't have space to discuss these developments here. We have no principled objection to deploying tools more sophisticated than two-tier MC-tests to appraise concept possession, if their effectiveness is attested.
[15] For further examples see a comprehensive list in Gurel et al. (2015: 996).

[*The correct answer may not be in this list. Your can write your own answer below.*]
G.   …   Therefore,   …

A pupil passes SHADOW* if and only if she selects both *B* and *F*. A pupil passes BEACH* just in case she selects *B* and writes in *G* the following statement (or one equivalent):

Its density is higher than water's. Therefore, it will sink in the water.

Suppose we use SHADOW* and BEACH* to assess pupils' concept possession. SHADOW*—in contrast to SHADOW—will allow us to conclude that Claire doesn't master the concept of light, for Claire will very probably *fail* SHADOW*. In fact imagine that Claire sits SHADOW* and selects *B* to answer question (1) because she believes that light alone can produce a shadow only when it is cast by a tree. Given her belief, Claire would very likely answer (2) by filling in the blank *distractor H* with the statement 'Because the shadow is cast by a tree' or an equivalent phrase. Since the correct response to (2) is *F*, Claire won't pass the test.

Let's turn to BEACH*. Recall that the problematic case involving BEACH doesn't depend on Claire's having any misconception. In this case Claire fills in details of the scenario in BEACH's stem in an unexpected way, and this leads her to select the distractor *C* rather than the key *B*. The problem is that Claire's selecting *C* doesn't indicate, in these particular circumstances, that she doesn't master the concept of sinking. Now suppose Claire takes BEACH* instead of BEACH, and that she selects *C* to answer (1) for the reasons just explained. Since the correct answer to (1) is *B*, Claire will "fail" BEACH*. Yet Claire's response to (2) will suggest that she may nevertheless master the notion of sinking. Since Claire selects *C* because she imagines that Carl fills the balloon with sand, Claire will probably answer (2) by writing in the blank distractor *G* the phrase 'It is full of sand. Therefore, it will sink in the water', or a similar phrase. This will illuminate why Claire has selected *C* and all at once show that Claire's reasoning is compatible with her indorsing the

inferences that define the notion of sinking. The question whether or not Claire actually masters the notion of sinking could at this point be settled through, for example, an oral exam.

Other problematic cases of assessment of concept possession similar to those involving SHADOW and BEACH can probably be resolved—or at least contained—by resorting to two-tier MC-tests. Test items of this type are nowadays occasionally deployed in primary and secondary education. For example, some of them usually feature in the National Curriculum Assessment (SATs) carried out in primary school in England, and in the OECD/PISA tests. We suggest that two-trier MC-tests should systematically be deployed whenever pupils' concept mastery is appraised through MC-tests.

7. Conclusion

We have focused on Brandom's inferentialism, according to which concepts are inferentially articulated entities. We have argued that the assessment of concept mastery at school based on MC-tests might prove unreliable because these quizzes don't check the inferences that pupils must draw or the reasons they could adduce to master a concept. We have proposed that MC-tests would be more reliable if they enabled pupils to make the reasons of their answers or the inferences they draw explicit. We have suggested that two-tier MC-tests might fulfil this role.

References

Allen, M. 2010. *Misconceptions in Primary Science*. Maidenhead: Open University Press.

Boghossian, P. 1996. 'Analyticity Reconsidered'. *Nous* 30: 360-391.

Brandom, R. 1994. *Making It explicit*. Cambridge, MA: Harvard University Press.

Brandom, R. 2000. *Articulating Reasons*. Cambridge, MA: Harvard University Press.

Byrne, R. 2005. *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA, MIT Press

Derry, J. 2013. 'Can Inferentialism Contribute to Social Epistemology?' *Journal of Philosophy of Education* 47: 222-235.

Derry, J. 2017. 'An introduction to Inferentialism in mathematics education'. *Mathematics Education Research Journal* (forthcoming). Published online at: https://link.springer.com/article/10.1007/s13394-017-0193-7

Devitt, M. 1981. *Designation*. New York: Columbia University Press.

Duit, R. and D. F. Treagust. (2003). 'Conceptual change: a powerful framework for improving science teaching and learning'. *International Journal of Science Education* 25: 671-688.

Dunlop, L., Compton, K., Clarke, L. and V. McKelvey-Martin. (2015).'Child-led enquiry in primary science'. *Education 3-13* 43: 462-481.

Etchemendy, J. (1990) *The Concept of Logical Consequence*. Cambridge, Mass., Harvard University Press.

Fodor, J. (1998) *Concepts*. Oxford: Clarendon Press.

Gurel, D. K, Eryilmaz, A. and L. C. McDermott. 2015. 'A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science'. *Eurasia Journal of Mathematics, Science & Technology Education* 11: 989-1008

Jorgensen, A. 2008. 'Understanding as Endorsing an Inference'. *Polish Journal of Philosophy* 2: 35-54l.

Haladyna, T. 2004. *Developing and Validating Multiple-Choice Test Items* (3rd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Hempel, C. 1988. 'Provisoes: A Problem concerning the Inferential Function of Scientific Theories', *Erkenntnis* 28: 147-164.

Lange, M. 1993. 'Natural Laws and the Problem of Provisos'. *Erkenntnis* 38: 233-248.

Murzi, J. and F. Steinberger 2017 'Inferentialism'. In B. Hale, A. Miller and C. Wright (eds.), *A Companion to the Philosophy of Language*, second edition, pp. 197-224. Oxford: Blackwell.

Peregrin, J. 2009. 'Inferentialism and the Compositionality of Meaning'. *International Review of Pragmatics* 1: 154–181.

Peregrin, J. 2012. 'Inferentialism and the Normativity of Meaning'. *Philosophia* 40: 75-97.

Rollnick, M. and P. P. Mahooana. 1999. 'A quick and effective way of diagnosing student difficulties: two tier from simple multiple-choice questions'. *South African Journal of Chemistry* 52: 161-164.

Sellars, W. 1949. 'Language, Rules and Behavior'. In S. Hook (ed.), *John Dewey: Philosopher of Science and Freedom*, pp. 289-315. New York: Dial Press.

Sellars, W. 1953. 'Inference and Meaning'. *Mind* 62: 313-338.

Sellars, W. 1974. 'Meaning as functional classification'. *Synthese* 27: 417–437.

Simkin, M. G. and W. L. Kuechler. 2004. 'Multiple-Choice Tests and Student Understanding: What Is the Connection?' *Decision Science* 3: 73-98.

Tamir, P. 1971. 'An alternative approach to the construction of multiple-choice test items'. *Journal of Biological Education* 5: 305-307.

Tamir, P. 1989. 'Some issues related to the use of justifications to multiple-choice answers'. *Journal of Biological Education* 23: 285-292

Tamir, P. 1990. 'Justifying the selection of answers in multiple-choice items'. *International Journal of Science Education* 12: 563-573.

Tsai, C. C. and C. Chou. 2002. 'Diagnosing students' alternative conceptions in science'. *Journal of Computer Assisted Learning* 18: 157-165.

Wang, J. R. 2004. 'Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system'. *International Journal of Science and Mathematics Education* 2: 131-157.