

A High Degree of Difficulty: Assessing Higher-order Thinking in Cross-curricular Tests with Multiple Choice Items

This paper reviews:

- definitions of higher-order thinking and reservations about the concept;
- the ways the multiple choice item format is inimical to testing higher-order thinking;
- how some carefully framed multiple choice can stretch the limits of the format to test higher-order interpretive reasoning, critical thinking and plausible reasoning;
- the way cross-curricular tests are particularly appropriate for assessing higher-order thinking.

The paper discusses:

- how cross-curricular tests contrast with other generic skills tests;
- a set of criteria for judging the quality of test items in generic skills tests;
- examples of multiple choice items that assess higher-order thinking.

The items examined will be drawn from the distinctive kind of cross-curricular tests produced by the Australian Council for Educational Research since the 1960s.

Doug McCurry
Senior Research Fellow
Australian Council of Educational Research
19 Prospect Hill Road
Camberwell
Victoria
Australia
0392775581
mccurry@acer.edu.au

Higher-order thinking is commonly typified as the three top three levels (Analysing, Evaluating, Creating) of Bloom's Revised Taxonomy (D.Anderson 2001). At the higher levels of thinking it is said that students are involved in designing, constructing, planning, producing, inventing, checking, hypothesising, critiquing, experimenting, judging, comparing, organising, deconstructing, interrogating and finding. In a typical statement, Kurwongbah State School in Queensland shows the usefulness of the term higher-order thinking by saying:

Higher-order thinking is not about regurgitation of information, it is not about rote learning or simple remembering or recall of facts. It is about engaging students at the highest levels of thinking to foster exciting learning environments where students become creators of new ideas, analysers of information and generators of knowledge.

In *Education and Learning to Think* Lauren Resnick (Resnick 1987) characterised higher-order thinking (HOT) as complex and non-algorithmic thinking involving:

- multiple solutions;
- nuanced judgement and interpretation;
- the application of multiple criteria;
- uncertainty;
- self-regulation of the thinking process;
- imposing meaning, finding structure in apparent disorder; and
- effort.

Mathew Lipman used the term HOT to mean 'conceptually rich, coherently organised and persistently exploratory' thinking that is 'critical, creative and caring' (Lipman 1991). Laurance Splitter considered the notion of HOT in *Teaching for Better Thinking* (Splitter 1995). He argued that the 'lower-order/higher-order dichotomy had passed its 'use by' date.' Lauren Resnick has also expressed concerns about the term HOT:

The most important single message of modern research on the nature of thinking is that the kinds of activities traditionally associated with thinking are not limited to advanced levels of development. Instead, these activities are an intimate part of even elementary levels of reading, mathematics, and other branches of learning - when learning is proceeding well. In fact, the term 'higher order' skills is probably itself fundamentally misleading, for it suggests that another set of skills, presumably called 'lower order,' needs to come first. This assumption - that there is a sequence from lower level activities that do not require much independent thinking or judgment to higher level ones that do - colors much educational theory and practice.

Even though there are reasons for concern about the term 'higher-order', the notion of HOT is valuable in that it encourages us to think about good thinking.

This paper is concerned with ways in which HOT can be assessed in standardised multiple choice question (MCQ) tests. HOT as envisaged by Resnick would not be easy to assess in any formal test, let alone in MCQ. It is much easier to assess HOT in short answer questions (SAQ) and writing tests which involve human judgement than in MCQ. But it should be noted that some SAQ are as closed and algorithmic as most MCQ. And some writing tests aim to be and are strictly convergent thinking. The Writing Sample of the Medical College Admission Test (MCAT), for instance, constrains the thinking of candidates within the following rigid restrictions.

Write a unified essay in which you perform the following tasks. Explain what you think the above statement means. Describe a specific situation in which a politician should take into account the beliefs of constituents when voting. Discuss what you think determines whether politicians should vote according to their beliefs or those of their constituents.

One can only wonder why a test for a very able candidature would use such a limited writing test. Such a writing test significantly reduces the cognitive challenge (and opportunity) involved in producing impromptu writing.

With Resnick's notion of HOT in mind, I am going to consider whether and how MCQ can test complex, non-algorithmic thinking involving:

- uncertainty and multiple solutions;
- nuanced judgement and interpretation; and
- imposing meaning and finding structure in apparent disorder.

I am going to argue that MCQ can be used to test HOT, but that it takes subtle use of the narrow MCQ format to do so. I am also going to argue that cross-curricular generic skills tests offer a special opportunity for developing HOT test items.

What is a cross-curricular test?

A cross-curricular (CC) test aims to assess the kinds of thinking that underpins general education. It does not test the knowledge or particular skills of a subject or area. A CC test is an assessment of generic skills or developed abilities test rather than an achievement test or examination. A CC test aims to minimise the importance of prior knowledge (they draw on general knowledge) by giving candidates unfamiliar stimulus materials to think about. Rather than testing what candidates 'know' when they come into the testing venue, a CC test assesses the ability of the candidate to reason about the material they are offered by the test.

A CC test contrasts with psychometric tests that test psychological constructs like verbal, quantitative and abstract reasoning. Such psychometric tests offer to assess basic skills. Verbal ability is understood as vocabulary knowledge and reading. Quantitative ability is understood as low verbal numerical processing. Abstract reasoning is understood as non-verbal, logico-deductive reasoning with symbols. In the name of efficiency, psychometric tests use different versions of the same item 10, 20, 30 times in a row. Leaving aside the tediousness of such repetition for the candidate, such psychometric items make no claim to having any meaning for candidates in the first place. CC tests, on the other hand, present candidates with stimulus to understand and construct meaning from. A CC test aims to be and should feel like real thinking to a candidate.

Perhaps I can make the distinctive nature of CC tests clearer by comparing them with a well known test that does not and cannot claim to be 'cross-curricular'.

The SAT Reasoning test

The Scholastic Aptitude Test developed by the Educational Testing Service for the College Board began in 1926 as a psychometric test including sections of definitions, arithmetic, classification, artificial language, antonyms, number series, analogies, logical inference, and paragraph reading. Over time it moved away from such purely psychometric constructs. By 1990 controversy about the notion of aptitude and intelligence testing meant the name of the test was changed to Scholastic Assessment Test. In 1994 the acronym SAT was changed to a word without any particular meaning. It is now named the SAT Reasoning Test without it seems any explanation of the meaning of the word SAT.

The quantitative reasoning section of the SAT is not know as the Mathematics sections. It aims to be 'straight maths'. The verbal reasoning section is now called Critical Reading (the vocabulary testing and the egregious verbal analogies have been dropped) with most of the emphasis is placed on reading of short and longer passages of prose. A writing section has recently been added.

A Tradition of Cross-curricular Testing in Australia

CC testing began in Australia in 1964 when the Australian Council for Educational Research (ACER) was requested by the Australian Government to develop an examination for awarding scholarships to senior secondary students. In the Commonwealth Senior Secondary Examination (CCSE) ACER used the procedures of standardised testing to develop with it called a 'new kind of testing' (Whitford 1966).

While the C.S.S.E. papers are quite different from traditional general ability or intelligence tests, it will be clear that they are also unlike either the school examinations or the public examinations with which students are likely to be familiar.

One obvious difference from the school or public examination is that the C.S.S.E. papers are not based on a specific course of study which all the examinees have followed, and which they have been deliberately taught. Instead, each paper attempts to test abilities which those at A.C.E.R. concerned with the examination believe will be developed by good teaching and interested learning, whatever the specific course of study. They are, we think, important abilities which are durable, and which should, if well developed, lead to success in later academic studies.

ACER recommended that the tests focus on the central areas of secondary education through four papers: Written Expression, Quantitative Thinking, Comprehension and Interpretation (Sciences), and Comprehension and Interpretation (Humanities).. The Director of ACER at the time, W. C. Radford, described the tests in the following way.

The abilities which they would test are those which should be developed in any good broadly based course of study, irrespective of its specific content . . . They would test, as a whole, the range of abilities which we think a soundly educated child should have at this level, i.e. [the child] should have a good background in literature, social studies, mathematics, and science (a background which has produced abilities which are not restricted to knowing or applying a particular content of a particular course, and which are transferable to new and different courses), should be able to write well, and should know something of the mechanics of expression. These are abilities which should be developed in any kind of school.

Radford's comments express of a new aim for general ability testing and a new and different relationship between ability testing and the school curriculum. The CSSE test developed into the

Tertiary Education Entry Project (TEEP), which in turn developed into the Australian Scholastic Aptitude Test (ASAT) which was first administered in 1974 (Connell 1980). In spite of the similarity in names, the ASAT was quite unlike the SAT. The ASAT aimed to test the kinds of thinking that underpins the school curriculum with two sections called Reasoning in Mathematics and Science and Reasoning in Humanities, Arts and Social Sciences.

The CSSE and the ASAT shaped the distinctive approach to cross-curricular testing taken by ACER in the 1970s and continued into such current tests as the Co-operative Scholarship Testing Program (CSTP), The Special Tertiary Admissions Test (STAT), The Australian Law Schools Entrance Test (ALSET), the Graduate Medical Entry Test (GAMSAT). This tradition of testing also evolved into the General Achievement Test (GAT, <http://www.vcaa.vic.edu.au/vce/exams/gat/index.html>) currently used for various kinds of educational monitoring in the state of Victoria, and the ACT Scaling Test (AST, http://www.bsss.act.edu.au/year_11_and_12/act_scaling_test) used as the external moderating device for adjusting school assessment in the Australian Capital Territory. This tradition of cross-curricular testing also shaped the development of the Queensland Core Skills Test (QCST, <http://www.qsa.qld.edu.au/assessment>), although there some significant difference between the foundation of that test and the ACER's cross-curricular tests that will be explained below.

The current incarnations of this tradition of CC testing used in government assessment programs involve an integrated paper of between 70 or 80 MCQ items undertaken in 120 to 180 minutes. The single paper is developed within two over-arching categories: Reasoning in Mathematics, Science and Technology (MST) and Reasoning in the Humanities, Arts and Social Sciences (HASS). Both the GAT and the AST have writing tests, but they differ significantly, and the AST also has a short answer question component.

The AST specification describes the test as 'designed to measure, in the main, the abilities of comprehension, interpretation and reasoning, across as many curriculum areas as possible, with a level of conceptualisation and difficulty appropriate to the final year of secondary schooling'. It further defines the content of the test in terms of subject or curriculum area, stimulus material and skills or abilities.

The AST aims to be a broad and balanced test of thinking skills. Rather than specific kinds of items that are slotted into different version of the test, each AST is in some sense unique and the construction of a form of the test involves subtle assessment of the breadth and balance of a proposed set of items. What is considered in constructing a broad and balance version of a CC test is a complicated story. How the tests forms are balanced in terms of curriculum area and stimulus material is straight-forward, but achieving balance in terms of skills and abilities is a matter of subtle judgment.

It is worth noting that the processes for developing items for the AST and the GAT at ACER differ from the gathering of items for most generic skills tests. The AST and GAT items are developed by full-time and experienced test writers. Professional test writers are needed to develop the kind of items used in tests like the AST and GAT. This work cannot be done by casual contributors. ACER has hardly ever been able to get material of anything like satisfactory quality from external writers. The cross-curricular tests ACER produces depend on the on-going collaboration of full-time item writing specialists. These writers are continually working on the kinds of thinking that are assessed in the MST and HASS components of our tests.

The QCST was developed out of the ASAT program, but it explicitly rejected the curriculum related nature of the MST and HASS constructs. It was to test the common curriculum elements (CCE) shown in (Figure 1). The CCE were developed from an examination of curriculum documents and an extensive consultation with teachers in all areas of the Years 11 and 12 curriculum in Queensland. Terms were included in the list if more than two subject areas identified it as indicating that an activity took place in their subject. The dominance of a Bloom-like terminology in educational discourse can be seen in the CCE. It was intended that the CCE would produce the broadest possible test. Assessing whether the QCST is broader than the AST is not a simple matter.

Rather than discussing the ways in which breadth and balance are considered in constructing a form of the AST or GAT, I am going to discuss some of the distinctive items testing HOT that are found in these CC tests. I am offering these examples to you as superior items that add to the breadth and balance of ACER's CC tests. Such items are not found in other generic skills tests because they could not be included in the constructs tested. I will limit my discussion to HASS items that elicit HOT by testing of plausible and interpretative reasoning.

Testing the Higher-order Thinking that Underpins the Humanities, Arts and Social Sciences

In my view superior MCQ are authentic tests of real thinking rather than artificial test gymnastics, like verbal analogies or number series items. Items are HOT to the extent that they test complex thinking rather than merely requiring difficult or complicated information processing. Such items ask the crucial and central questions about rich stimulus material. They attempt to minimise the testing of

knowledge, and they attempt to test the ability of candidates to understand and learn from unfamiliar stimulus material. They are culture rich without being tests of cultural knowledge. They are substantially conceptual rather than verbal, and they commonly use diagrammatic and visual material for stimulus to partially mitigate the limitation of a reading based test. The items even aim to be interesting and engaging for candidates.

HASS reasoning tests skills in the interpretation and understanding of ideas in social and cultural contexts. Different kinds of text are used as stimulus, including passages of personal, imaginative, expository and argumentative writing. Although most of the stimulus materials in the HASS section of CC tests are in the form of written passages, some units may present ideas and information in pictorial, diagrammatic or tabular form. Materials deal with a range of academic and public issues, with an emphasis on socio-cultural, personal and interpersonal topics.

HASS questions demand varying degrees of complex verbal processing and conceptual thinking, logical and plausible reasoning, and objective and subjective thinking. Some questions that emphasise understanding involve the recognition of explicit and implicit meanings through close reading of words and phrases and global interpretations of texts. Some questions that emphasise plausible reasoning involve interrelating, elaborating and extending concepts and ideas, and drawing conclusions. Other questions that emphasise critical thinking require candidates to make discriminations and judgments in the realm of plausible reasoning. In summary HASS reasoning involves:

- reading and interpreting written text, pictures, diagrams and numerical data;
- analysing and understanding concepts and the relationship between different concepts;
- grasping the implications of propositions and ideas;
- seeing the direction and upshot of arguments; and
- identifying the way different arguments related to each other.

The validity of CC tests depends on being authentic, broad and balanced rather than a narrow and artificial test of reasoning. While HASS items usually involve a good deal of reading, the items aim to be more than a test of reading test. While some HASS items might be described as primarily reading comprehension, other items aim to minimise verbal interpretation and emphasise conceptual thinking.

Plausible Reasoning in HASS Items

A key feature of HASS is the aim to test different kinds of reasoning. Some HASS items test the kinds of logico-deductive reasoning that are typically tested in the MST items. Answers to MST questions can almost always be explained with a clear and indisputable process of reasoning. Candidates may not come to conclusions about MST questions by such a process (they may also have insights, jump steps and recognise rather than deduce conclusions), but the answers to MST questions can be readily explicated in a logico-deductive and linear fashion. Some HASS items are of this kind, but other items involve different kinds of reasoning. Some HASS reasoning involves analogical or associative reasoning processes, and some of it involves judgements of relative significance or value. Some of HASS items involve the recognition of dialectical relationships that is the fundamental characteristic of critical thinking (Facione 1990).

The notion of plausible reasoning is a key concept for understanding HASS as developed by ACER. Plausible reasoning involves thinking that is analogical and associative rather than logico-deductive. The mathematician G. Polya distinguished between demonstrative reasoning and plausible reasoning:

We secure our knowledge by demonstrative reasoning, but we support our conjectures by plausible reasoning. A mathematical proof is demonstrative reasoning, but the inductive evidence of the physicist, the circumstantial evidence of the lawyer, the documentary evidence of the historian, and the statistical evidence of the economist belong to plausible reasoning' (Polya 1954).

The kind of plausible reasoning involved in HOT HASS items contrasts with most of the reasoning involved in other kinds of multiple choice testing. Such items are hard to write, and they are hard to make discriminate as other multiple choice items do, but such plausible reasoning items are a valuable diversification of what is usually the overwhelmingly logico-deductive and linear nature of most multiple choice testing.

While in some sense all the HASS questions involve logical reasoning, candidates are also required in some items to make inferences and interpretations that are not logically irrefragable. These plausible reasoning items require candidates to recognise strong inferences and interpretations from implausible claims or inferences. Judgements of this kind are not logically necessary or objectively and/or empirically true. Such items deal in some sense with subjective matters and value judgements, and in doing such items candidates have to recognise plausible interpretations from implausible ones.

The following HASS items are discussed in terms of the kind of HOT they involve, and the way they contribute to an authentic, broad and balanced test of reasoning.

Some Examples of HOT HASS Items

The use of cartoon as stimulus is a distinctive part of CC testing at ACER tests. The thinking involved in interpreting cartoons is fundamental to achieving the aims of our CC testing. Cartoons are not to be found in the SAT, and there is little of the plausible reasoning that typifies HOT items in a reading test like the SAT. Test items on cartoons are the epitome of the interpretive and plausible reasoning that underpins much HASS thinking.

In Example 1 interpreting the cartoon depends on recognising the three geometric shapes and the TV and interpreting the relationship between these things. The MCQ options offer candidates four apparently or superficially plausible interpretations, but the key is a very strong interpretation, and the distractors are not really plausible.

It should be noted that the key to this question is not a 'best answer', in the sense intended in some MCQ items where some distractors are judged to be merely or moderately plausible but not as good as the key. An interpretive question like this has a very strong interpretation as the key, but the distractors are not less good answers or vaguely plausible interpretations. We require distractors to be implausible, and argue that they can be eliminated on logical grounds.

In Example 1 option C is wrong because it does not deal with the geometric shapes and the TV. Option A offers an interpretation of the geometric shapes and the TV, but it does not explain why TV is taken to be symbolic of 'everything'. The TV shape in the cartoon symbolises TV. B interprets the geometric shapes as 'fundamentals' and contrasts TV with them, but it does not offer a meaningful interpretation of the whole. D is the key because it explains why TV has been brought together with the fundamental geometric shapes.

Example 2 offers four interpretations of missiles as a technology in the cartoon. The primitive figure swinging the missile as a battle axe is neither 'sophisticated', 'precisely controlled' nor 'particularly dangerous'. As stated in the key, the cartoon suggest that missiles are not more 'sophisticated' weapons than battle axes.

Example 3 asks interpretive questions about a poem and the painting on which the poem is based. By putting the two texts together the difficult job of pinning down a powerful interpretation in such material is reduced.

The questions ask about the poet's interpretation of the image. O'Callaghan tells us that the nighthawks are 'driven' to the diner, and, although they sit 'side by side', they are trying not to count the cigarettes, coffee and the time passing rather than communicating. Hence the key to the first item is 'vacant and uncommunicative'.

The second item asks for an interpretation of the diner itself, and the key picks up O'Callaghan's description of it as an illuminated fish tank which suggests 'a kind of exposure' as the key.

Example 4 is one of the ways in which a reading based CC test can attempt to minimise reading and maximise conceptual thinking. Five theories that explain and justify belief in human rights are simply and explicitly defined, and the questions ask about the nature of the theories and the relationship between them. Candidates are asked about which theory would be best described as 'moral', which two theories have most in common, and what it means to treat humans rights as a contract.

Example 5 is a comparatively low verbal example of critical thinking. Candidates have to decide how a particular statement relates to a proposition under debate. They also have to decide how a particular statement can be most appropriately rebutted.

In my view items such as these test interpretive ability, critical thinking and plausible reasoning. With Resnick's definition of HOT in mind, I would argue that such MCQ test complex, non-algorithmic thinking involving:

- the possibility of multiple solutions;
- nuanced judgement and interpretation; and
- finding structure in apparent disorder so as to construct meaning.

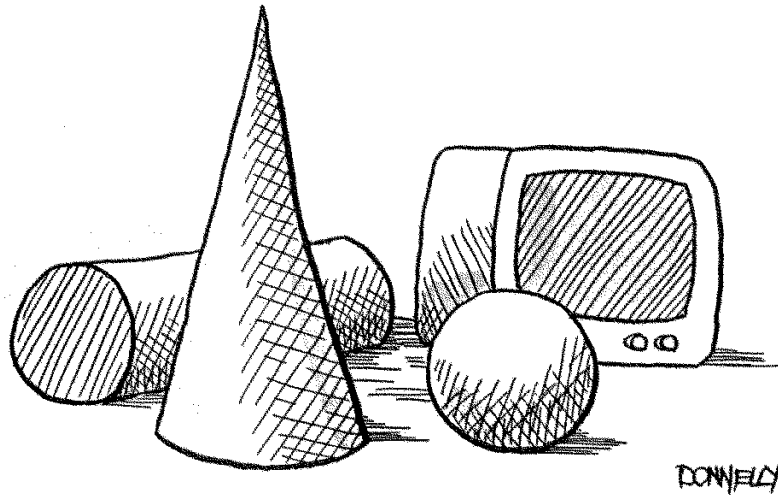
At their best, such HASS MCQ items are HOT because they test complex thinking.

Figure 1 The Common Curriculum Elements of the Queensland Senior School

- 1 Recognising letters, words and other symbols
- 2 Finding material in an indexed collection
- 3 Recalling/remembering
- 4 Interpreting the meaning of words or other symbols
- 5 Interpreting the meaning of pictures/illustrations
- 6 Interpreting the meaning of tables or diagrams or maps or graphs
- 7 Translating from one form to another
- 9 Using correct spelling, punctuation, grammar
- 10 Using vocabulary appropriate to a context
- 11 Summarising/condensing written text
- 12 Compiling lists/statistics
- 13 Recording/noting data
- 14 Compiling results in a tabular form
- 15 Graphing
- 16 Calculating with or without calculators
- 17 Estimating numerical magnitude
- 18 Approximating a numerical value
- 19 Substituting in formulae
- 20 Setting out/presenting/arranging/displaying
- 21 Structuring/organising extended written text
- 22 Structuring/organising a mathematical argument
- 26 Explaining to others
- 27 Expounding a viewpoint
- 28 Empathising
- 29 Comparing, contrasting
- 30 Classifying
- 31 Interrelating ideas/themes/issues
- 32 Reaching a conclusion which is necessarily true provided a given set of assumptions is true
- 33 Reaching a conclusion which is consistent with a given set of assumptions
- 34 Inserting an intermediate between members of a series
- 34 Inserting an intermediate between members of a series
- 35 Extrapolating
- 36 Applying strategies to trial and test ideas and procedures
- 37 Applying a progression of steps to achieve the required answer
- 38 Generalising from information
- 41 Hypothesising
- 42 Criticising
- 43 Analysing
- 44 Synthesising
- 45 Judging/evaluating
- 46 Creating/composing/devising
- 48 Justifying
- 49 Perceiving patterns
- 50 Graphicalising
- 51 Identifying shapes in two and three dimensions
- 52 Searching and locating items/information
- 53 Observing systematically
- 55 Gesturing
- 57 Manipulating/operating/using equipment
- 60 Sketching/drawing

http://www.qsa.qld.edu.au/downloads/assessment/qcs_elements.pdf

Example 1 Geometric shapes and TV



The drawing above suggests that

- A geometric shapes are behind everything, even television.
- B television contrasts with other fundamentals.
- C geometric shapes are infinitely flexible.
- D television has become fundamental. **

Example 2 Missile technology



This illustration suggests that missiles are

- A a sophisticated technology.
- B an unsophisticated technology. ***
- C a precisely controlled technology.
- D a particularly dangerous technology.

Example 3 Nighthawks

The following poem was written by Julie O'Callaghan in response to **Nighthawks** painted by Edward Hopper in 1942.

Nighthawks

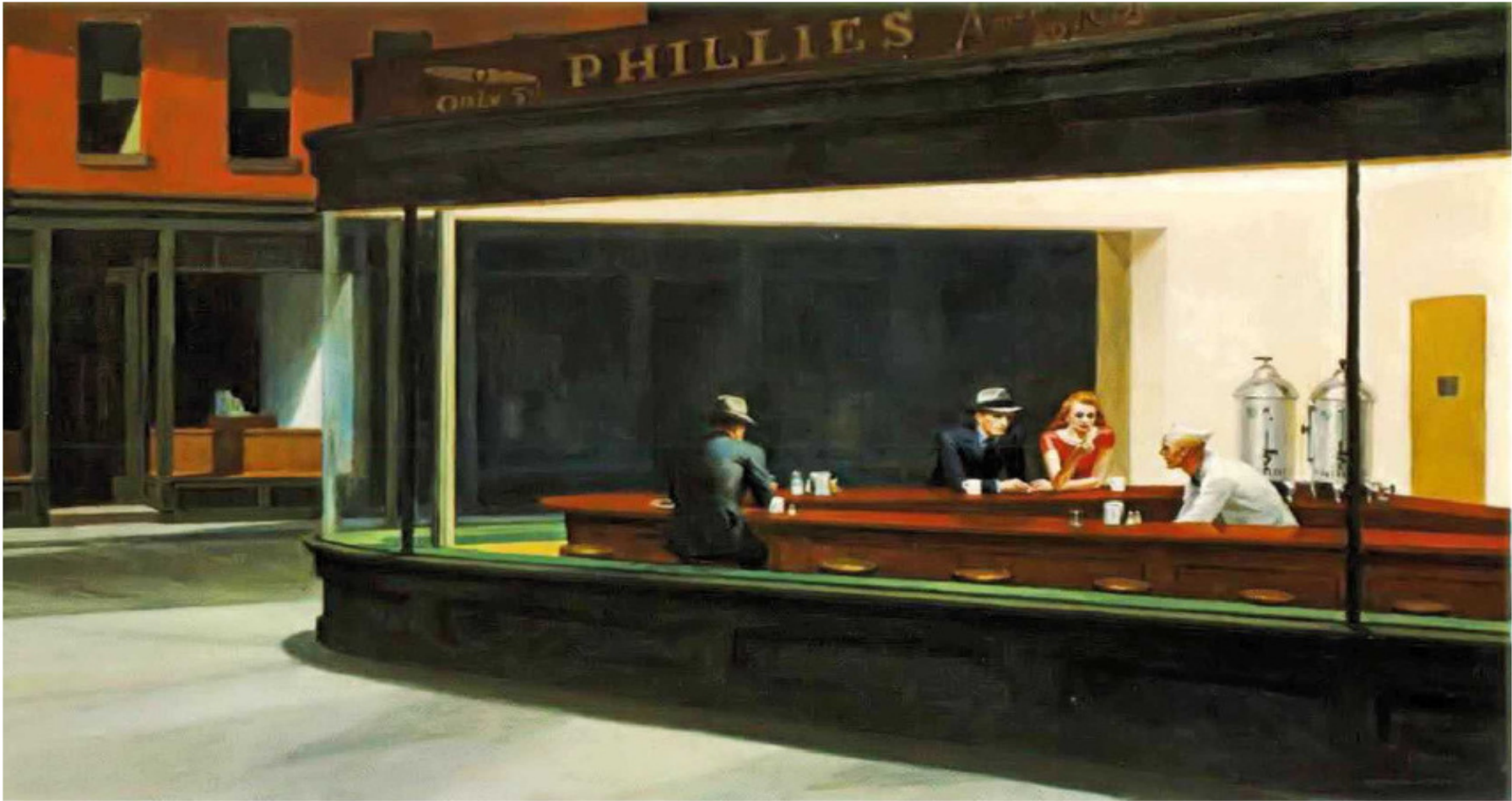
The heat and the dark
drive us from apartments
down empty streets
to the all night diner
where the fluorescent lights
illuminate us like tropical fish
in a fish tank.
We sit side by side
listening to glasses clink,
the waiter whistling,
and stare at the concrete outside.
Not looking at our watches
or counting the cigarettes
and cups of coffee.

O'Callaghan's poem presents the customers in the diner as

relaxed and relieved.
anxious and nervous.
vigilant and expectant.
vacant and uncommunicative. ***

O'Callaghan gives an impression of the diner as

a soothing retreat.
a kind of exposure. ***
potentially explosive.
glamorous and stylish.



Example 4 Theories of Human Rights

Several approaches have been used to explain and justify the belief in human rights.

Biological theories	see human rights as arising from the reproductive advantage of human social behaviour, and based on concern for other people in the race to continue the species.
Social contract theories	claim that individuals in a society accept rules in exchange for security and economic advantage.
Natural law theories	base human rights on the view that there is a natural moral order of objectively valid prescriptions.
Supernatural law theories	see human rights as based on the law of God as presented in such religious texts as the Bible and Qur'an.
Humanist theories	see human rights as based on universally applicable values.
Interest theories	claim that human rights are justifiable on the grounds of their value in creating the necessary conditions for human well-being. Some interest-theorists also justify human rights on grounds of self-interest (rather than concern for others). Respecting the rights of others ensures that one's own will be protected.

Which one of the following is best described as a practical rather than a moral kind of theory?

- A Interest theories ***
- B Humanist theories
- C Natural law theories
- D Supernatural law theories

Describing human rights as a 'contract' suggests that human rights

- A cannot be cancelled.
- B have a moral but not a legal force.
- C have a moral force above the law.
- D are an agreement that has the force of law. ***

Which of the following pairs of human rights theories have the most in common?

- A Humanist theories and Biological theories
- B Biological theories and Supernatural law theories
- C Interest theories and Biological theories ***
- D Supernatural theories and Interest theories

Example 5 The Violence of TV Debate

Violent TV and videos contribute to real-life violence.

For each of the questions you are to choose the alternative (A – D) that most appropriately describes the relationship of the statement to the topic of the debate.

The statement:

- A is most likely part of the debate **for** the topic.
- B is most likely part of the debate **against** the topic.
- C could possibly be part of the debate **for or against** the topic.
- D is **not relevant** to either the debate for or against the topic.

Question 12

People easily separate fantasy from reality.

Question 13

Violence is never an acceptable course of action.

Question 14

Violent entertainment desensitises viewers.

Question 15

The accessibility and the graphic impact of TV make it an especially powerful medium.

References

- Connell, W. F. (1980). The Australian Council for Educational Research 1930-80. Melbourne, Australia.
- D.Anderson, L. W. K. (2001). A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. New York, Addison Wesley Longman.
- Facione, P. A. (1990). "Critical Thinking: a Statement of Expert Consensus for Purposes of Educational Assessment and Instruction."
- Lipman, M. (1991). Thinking in Education. New York, Cambridge University Press.
- Polya, G. (1954). Mathematics and plausible reasoning. Princeton, Princeton University Press.
- Resnick, L. B. (1987). Education and Learning to Think. Washington, DC, National Academy Press.
- Splitter, L. J. a. S., Ann M. (1995). Teaching for Better Thinking: The Classroom Community of Inquiry. Melbourne, Australia, The Australian Council for Educational Research Ltd.
- Whitford, T. M. (1966). Objectives Being Tested in the Commonwealth Secondary Scholarship Examination prepared in 1966. Melbourne, ACER.