Thinking Skills and Creativity 10 (2013) 250–264

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: http://www.elsevier.com/locate/tsc

Higher order thinking in chemistry curriculum and its assessment

Peter J. Fensham*, Alberto Bellocchi

Queensland University of Technology, Victoria Park Road, Brisbane, Queensland 4053, Australia

ARTICLE INFO

Article history:
Received 17 December 2012
Received in revised form 11 June 2013
Accepted 26 June 2013
Available online 12 July 2013

Keywords:
Chemistry assessment
Higher-order thinking in chemistry
Chemistry curriculum
PISA questions
Assessing higher-order thinking

ABSTRACT

Higher-order thinking has featured persistently in the reform agenda for science education. The intended curriculum in various countries sets out aspirational statements for the levels of higher-order thinking to be attained by students. This study reports the extent to which chemistry examinations from four Australian states align and facilitate the intended higher-order thinking skills stipulated in curriculum documents. Through content analysis, the curriculum goals were identified for each state and compared to the nature of question items in the corresponding examinations. Categories of higher-order thinking were adapted from the OECD's PISA Science test to analyze question items. There was considerable variation in the extent to which the examinations from the states supported the curriculum intent of developing and assessing higher-order thinking. Generally, examinations that used a marks-based system tended to emphasize lower-order thinking, with a greater distribution of marks allocated for lower-order thinking questions. Examinations associated with a criterion-referenced examination tended to award greater credit for higher-order thinking questions. The level of complexity of chemistry was another factor that limited the extent to which examination questions supported higher-order thinking. Implications from these findings are drawn for the authorities responsible for designing curriculum and assessment procedures and for teachers.

1. Introduction

The development of thinking skills, or more specifically higher order thinking (HOT) skills, in school science is a desirable educational goal that features regularly in educational reform agendas, science curriculum documents, and the science education literature (Gallagher, Hipkins, & Zohar, 2012). However, there are concerns internationally that dominant assessment practices in science focus on low order thinking skills and that this in turn encourages teachers to focus on pedagogies that emphasize rote-learning (e.g., Osborne & Dillon, 2008). This suggests that the assessment instruments associated with chemistry curricula play significant roles in supporting or hindering the intent for HOT. This paper is concerned with the degree to which HOT is expected in high school chemistry curriculum documents and to what extent examinations support or discourage HOT.

* Corresponding author. Tel.: +61 0731383338.
E-mail addresses: p.fensham@qut.edu.au (P.J. Fensham), alberto.bellocchi@qut.edu.au (A. Bellocchi).

1871-1871/$ – see front matter
http://dx.doi.org/10.1016/j.tsc.2013.06.003

2. Approaches to higher order thinking

Systematic research interest in HOT originated in the contribution of Bloom (1956) who suggested a hierarchy of intellectual skills based on six verbs. The lower three were *recall*, *comprehend* and *apply*, and the upper three were *analyze*, *synthesize*, and *evaluate*. Bloom's Taxonomy inspired one of us (Fensham, 1962) to begin his research in science education by reanalyzing the results of a set of university chemistry students he had taught. These students had undertaken an assessment test that required them to answer five questions (20 marks each) out of eight. Three of the questions were designed to require HOT based on Bloom's Taxonomy. The students who chose to answer these HOT questions had a bipolar score distribution compared with the more normal distributions for the other five questions. This turned out to be due to the HOT questions being avoided by high achievers and being attempted by low achievers who failed to recognize their difficulty. In the reanalysis of the overall student rankings, greater weighting was given to the scores on the HOT questions and this led to a considerable reordering of the ranked scores across the set of students.

Subsequent authors have developed different hierarchies from Bloom's Taxonomy to indicate the levels of reasoning or understanding in student responses to assessment instruments involving open-ended questions. The Structure of Observed Learning Outcomes (SOLO) taxonomy of Biggs and Collis (1982) has five levels of increasing complexity from pre-structural to extended abstract, that depend on how the student's response (for level 1) involves single or unrelated pieces of information or (for level 5) integrates numerous pieces of information and then applies the integrated information to new or untaught situations (level 5). The verbs in the presenting question can be indicative of the level of response expected. For instance, *analyze*, *apply*, *argue*, *compare/contrast*, *criticize*, *explain causes*, *relate*, and *justify* encourage level 3 and *create*, *formulate*, *generate*, *hypothesis*, *reflect*, and *theorize* encourage level 4. In addition to the use of verbs, the SOLO taxonomy also integrates the task complexity to the verb descriptors so that student responses that contain only single or unrelated pieces of information are deemed to be *unistructural* responses. Students' responses that integrate numerous pieces of information and then apply the integrated information to new or untaught situations are deemed as *abstract relational responses*.

Other taxonomies based on cognitive levels have been developed as tools to assist in the alignment of curriculum objectives with assessment and teaching practices (Webb, 1997, 2007). In this alignment procedure Depth of Knowledge in both the objectives and the assessments is indicated by four levels of mental processing: Level 1 (recall), Level 2 (skill/concept), Level 3 (strategic thinking) and Level 4 (extended thinking). The kinds of activities that students could perform to demonstrate competence at level 4 included "(a) developing and proving conjectures, (b) designing and conducting experiments, (c) making connections between a finding and related concepts and phenomena, (d) combining and synthesizing ideas into new concepts, and (e) critiquing experimental designs" (Webb, 2007, p. 13).

The analytical framework that informed our study is derived from the measures for scientific literacy used in the OECD's Programme for International Student Achievement (PISA) project. These assessment measures were employed with 15-year-old students and had characteristics that reflect each of the aforementioned taxonomies (i.e., Bloom's, SOLO and Webb's). In this PISA Science project, a clear distinction is made between knowing a science concept, principle, or procedure, and the active application of that knowledge to unfamiliar situations. This distinction is perhaps also implied in curriculum statements which commonly use the two words, "Knowledge and Understanding," as a learning objective. A set of test items in the PISA project was developed for each of three scientific competences: *explaining science phenomena, investigating science phenomena*, and *using scientific evidence.* The test items were intended to present different cognitive levels of application of science and technology in unfamiliar contexts. The items included a mixture of item modes including single multiple choice, complex multiple choice, and open constructed answers. In practice, the different degrees of item difficulty were determined post hoc by the percentages of students succeeding with an item. Six levels of difficulty were established with 2% of the very large student sample from the many participating countries achieving the highest level 6 items and 90+% achieving the lowest level 1 items (OECD, 2007, 2010). Examples of the prose descriptors used to differentiate between levels 1, 3, and 6 are shown in Table 1.

Clear differences can be seen for these three levels in terms of how the cognitive verbs (identify, explain, apply) interact with different amounts of information and the familiarity of their applications. This provides an operational definition that goes beyond the mere use of verbs as in Bloom's Taxonomy. For example, at level 6 the specification of "complex life

Table 1
PISA 2006 Proficiency Levels: using scientific evidence (OECD, 2007).

Level	Descriptor of student performance
6	At Level 6, students can consistently identify, explain and apply scientific knowledge and knowledge about science in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions. They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they demonstrate willingness to use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that centre on personal, social or global situations.
3	At Level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.
1	At Level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and that follow explicitly from given evidence.