Eurasian Journal of Physics and Chemistry Education

Eurasian J. Phys. & Chem. Educ. 5(2): 164-175, 2013

journal homepage: http://www.eurasianjournals.com/index.php/ejpce

Investigation of Questions in Science and Technology Textbooks in terms of Requirements of the Curriculum after Educational Reform in Turkey

Süleyman Yaman*

Ereğli Education Faculty, Bülent Ecevit University, Turkey

Received: 07 May 2013- Revised: 25 September 2013- Accepted: 27 September 2013

Abstract

In this study, the pupose is to investigate the questions in science and technology textbooks in terms of number of questions per unit in line with unit weight that indicates relative number of questions their cognitive levels and adopted evaluation approaches in the curriculum. In two forth and two fifth grade texbooks, 3474 questions were investigated by using which content analyses method. As a result of the study, differences in distribution of the questions both in one textbook and between the texbooks were detected based on comparing weight of unit and weight of corresponding question number within that unit. It was found that numbers of the questions on input and processing skills as cognitive level determinants was more than number of output skill questions and over the half of the questions were located at in-text position.

Keywords: Textbooks, science and technology curriculum, question types, cognitive level

Introduction

Various reform efforts have been made in the area of science education in Turkey since 1990 (Ayas, Çepni, Johnson & Turgut, 1997). Gür, Çelik and Özoğul (2012) stating that level of learning outcomes of studens in Turkey being very low, said that increasing level of education is necessary in Turkey. One of the prevention methods used to increase quality of education is to make serious changes in teaching programs. Majority of these reforms have taken into account how further development of students in science education could be provided, and how their achievements in science could be increased (AAAS, 2001; NRC, 1996). As the case in other countries like US and UK, authorities of the Turkish education system also reformed the science education curricula, since achievements of students in both elementary and secondary level science and math courses were lower than expected. Moreover, achievement levels of the students in these courses were influencing achievement levels in other courses (Berberoğlu, Kaptan & Kutlu, 2002).

Turkish science education curricula were changed for six times respectivelly in 1968, 1974, 1977, 1992, 2000, and 2005. However, none of the reforms could overcome the problem of low levels of achievement in science based on the students' contitinuous low performance in national as well as international examinations (Berberoğlu & Kalender, 2005). This persistence underachievement in science forces a fundamental drive for making another reform in science education curricula (MoNE, 2007). Current programs have been used in 2013. The effects of this change will occur in later years.

*Corresponding Author, Phone: +903723233870-126, Fax: +903723238693,
E-mail: syaman@karaelmas.edu.tr

ISSN: 1306-3049,

In fact, considering only the science curriculum as for the reason of low levels of student achievement, while not focusing on other factors, such as teachers, students, physical environment, materials, evaluation and course textbooks, would not provide a pertinent approach to reach a sustainable solution. Especially, science course textbooks are very important materials for a science curriculum, while they guide teachers and provide basic resources for students (Stake & Easley, 1978; Yager, 1996). On the other hand, using only science textbooks for teaching in science courses would be as problematic as not using any textbook (Armbruster & Ostartag, 1989). Colburn (2000) showed that over-dependence on science textbooks is a significant cause of low levels of achievement in science, and suggested that science textbooks should be used only as guides and secondary resources because science textbooks provides indirect knowledge on scientific topics. Colburn claimed that students should learn science by their own experiences and conduct researches for their primary resources. Newport (1965) explained the role of science textbooks on lower and higher levels of achievement and emphasized the importance of science textbooks for reaching curriculum objectives.

Being frequently used resources of instruction, characteristics of science textbooks are analyzed by some researchers, who emhasized the quantity and quality of questions in science textbooks as important characteristics that should be considered (Armbruster & Ostartag, 1989; Oakes & Saunders, 2004). Studies on the quality of questions have been conducted for a long time (Gall, 1970; Pizzini, Shepardson & Abell, 1992; Stern & Ahlgren, 2002). Brill and Yarden (2003), and Ogan-Bekiroglu, (2007) showed that textbook questions have been contributing to the general development of skills for lower-level-thinking rather than skills for higher-level-thinking. Meyer, Crummey, and Greer (1988) demonstrated the inappropriate distribution of science textbook questions based on grade levels. By focusing on a different aspect, Armbruster and Ostertag (1993) claimed that there are significant differences in the distribution of questions between social studies textbooks and science textbooks in terms of the purposes of questions. Yılmaz, Seçken, and Morgil (1998) investigated high school chemistry textbook questions based on the views of teachers and showed that questions were not competent enough for the achievement levels of students at their corresponding grade. In a different vein, Stern and Ahlgren (2002) stated that questions at the end of course materials did not develop the scientific literacy levels of students or at least provide partial improvement of scientific literacy. These studies presented ebove showed inappropriate distribution of questions within textbook based on matching between weight of the units and corresponding numbers of the questions per unit. For the location problem, Costa (1985) has provided a frame of classification in order to classify questions in terms of location and cognitive levels.

In this study Costa's classifi cation has been used because, as Dávila and Talanquer (2010) also stated, it is both in line with Bloom *et al.* (1956)'s classification and appears to be comprehensive. Costa (1985) has emphasized the relevance of the structure of questions in shaping cognitive thinking abilities and attitudes of students. Based on this assertion, questions have been classified in this study into three different types: (i) input, (ii) processing, and (iii) output questions. *Input questions* aim to lead students to recall their knowledge that is based on their prior experiences, long and short term memories. This type of questions target lower-level-thinking abilities that correspond to the "knowledge level" of Bloom's taxonomy (Dávila & Talanquer, 2010; Pizzini, Shepardson & Abell, 1992). These questions are to evaluate actions like description, completion, listing, selection, observation, telling, summarizing and recalling. Some examples of input questions are "What would be the word that explains best this picture?", "What are the names of children in tale" or "What do you feel about this event?" The second type of questions, namely, *processing questions* focus the meaning that students attribute to their knowledge. This type of questions corresponds to the “comprehension level” in Bloom’s taxonomy (Dávila & Talanquer, 2010). They focus on understanding causal relationships, making analysis, synthesis, comparisons, selections, classifications, as well as making analogies. Some examples of processing questions are “Why does matter melt?” and “What are the similarities between these two tools?” *Output questions*, which constitute the third type of questions, are about use of knowledge in new situations (Pizzini, Shepardson & Abell, 1992). These questions correspond to “synthesis” and “evaluation levels” in Bloom’s taxonomy (Dávila & Talanquer, 2010). Output questions concern skills of applying, imagination, planning, setting hypothesis, generalization and modelling. An example would be “What is the best solution for this problem?”, and “If we put freshwater fishes into sea water, what would happen?”

In the literature, there are studies using Costa’s classification to analyze questions in textbooks. Shepardson and Pizzini (1991) have also classified questions in science textbooks according to their purposes (based on Costa’s classification) into input, processing and output questions. They claimed that the majority of questions in science textbooks were input questions, and that the numbers of processing and output questions were not adequate. In the Turkish context, Kahveci (2010) analyzed science technology textbooks by using the same classification and found that the number of input and processing questions exceeded output questions.

Text book questions provide good means of feedback in checking whether curriculum objectives are attained. In that sense, the effectiveness and quality of these questions are directly related to the effectiveness of the curriculum itself. In the new Turkish science and technology curriculum, constructivist approach has been adopted. In line with this approach, alternative evaluation techniques have been emhasized rather than traditional measurement and evaluation techniques (Gelbal & Kelecioglu, 2007). Traditional measurement and evaluation techniques include multiple-choice items, true-false items, cloze, open-ended questions with short or long-answer essays, and matching. Alternative evaluation techniques include portfolio, concept maps, constructed grid, word association, project and perfomance evaluation (MoNE, 2005). Beyond alternative evaluation, the new curriculum also emphasized process and product evaluation together.

When use of science textbooks by majority of science and technology teachers is considered (Radcliffe, Caverly, Hand & Franke, 2008), analysis of science textbooks in terms of the quality and quantity of its questions by comparing actual situations to objectives and requirements of the curriculum will provide important feedback for its implementation. In that sense, the purpose of this study is to analyze the questions in fourth and fifth grade science and technology textbooks in terms of their quality and quantity. Based on the requirements of science and technology curriculum, the following research questions have been examined:

1. Is there any imbalance in the distribution of questions in terms of the weight of curriculum objectives per unit in different textbooks?
2. Is there any inconsistency in the distribution of different types of questions in terms of cognitive levels in different textbooks?
3. Is there any difference in types of questions across different textbooks?

Methodology

In this study, the quantitative content analysis technique was used in order to analyze textbook questions (Berelson, 1952; Cited by Rourke & Anderson, 2004). Contents of a document are analyzed by the frequency of particular categories of meaning in content analysis (Jupp, 2006). Categorical content analysis was utilized as a type of content analysis.

It is conducted by coding raw data and establishing categories based on associated data in the content. In the analysis, the examined text is initially partitioned into different sections, and then these sections are categorized by assigning them to related groups based on pre-determined standards (Tavşancıl & Aslan, 2001). In this study, content of the analysis is limited to questions in four science and technology textbooks; two textbooks for each grade levels (fourth and fifth grades). Textbooks examined in this study are made of the distribution for each province by the Ministry of National Education (MoNE).

Krippendorff (2003) explained that there are five basic processes of content analysis. These are determination of units of analysis, selection of sample, narrowing, meaning making and explaining. In the current study, science and technology textbooks were selected as basic units of analysis. Four textbooks were included in sample of the study. For narrowing the content, questions were analysis by considering Costa's taxonomy in terms of the purposes, types and rates of questions according to the number of curriculum objectives. During the process of analysis, common evaluation techniques were classified based on classifications of Anderson (1972), Armbruster and Ostertag (1993), and Haladyna (1992; 2004), whereas alternative evaluation techniques were classified according to the Turkish science and technology curriculum (MoNE, 2005). The data obtained were transferred into SPSS sheet. The data were statistically analyzed by taking into account variables that are stated in the research questions of this study. The analysis is divided into two sections of descriptive and comparative analyses. Consistency in number of questions regarding a unit in the science and technology textbooks and number of objectives in corresponding unit is also examined based on data of this study.

Selection of the Textbooks

The textbooks analyzed in this study were approved by the Turkish Ministry of Education. They were published in 2010 and were distributed to elementary schools in the same year. Purposive sampling was used in the selection of textbooks, considering their commonness and accessibility. Two of the textbooks were prepared by the MoNE, while the other two were prepared by private publication companies. They were approved and validated for a period of five years in Turkish elementary schools by MoNE.

The Turkish science and technology curriculum for fourth and fifth grades was in use since 2005. Until now, the curriculum was used without any revision. For higher grades, though, science and technology currriculum was applied first in 2006. In this study, specifically fourth and fifth grade textbooks were consulted because of their use for a longer time. Based on two possible publication authorities using the same curriculum, two textbooks by each authority were selected.

Data Collection and Analysis

In general, two different approaches are used for the analysis of textbooks. First one entails analysis of randomly selected samples from the contents of the textbooks, while the second one consist of the analysis of the textbooks as a whole (Wang, 1998). In this study, the second approach was chosen in order to increase coverage of units of analysis. In addition, two different field experts were consulted to decrease errors coming from researchers to increase the reliability. Interrater reliability of the scores was found as .94. After investigation of the agreement between researchers, problematic points were determined and then revisions were asked from the experts to reach consensus.

The analysis showed that the number of questions categorized as alternative evaluation techniques was very limited, so that all of them were grouped under a specific category, "alternative evaluation questions", while traditional evaluation questions were grouped into different categories. The obtained data were analyzed by using SPSS. In the analysis process, percent and frequency values were used in order to describe data. For example, percentages apearence of the question were computed within and across the groups and the relative percentages were calculated. Addionally, chi-square technique was employed to compare groups of evaluation approaches and the numbers of questions in the textbooks.

Results

The results of analyses focused on the questions in fourth and fifth grade textbooks by comparing the number of questions per units, in line with weigts of the units, locations of questions, cognitive levels of questions and their types were presented in this section. Table 1 presents rate of question number to total number of objectives in science and technology education curriculum.

Table 1. Rate of Question Number to Number of Objectives in Science and Technology Curriculum

Grade	Books	f	no	f / no
Fourth	MoNE4	648	178	3.48
Fourth	PRV4	922	178	5.10
Fifth	MoNE5	888	196	4.51
Fifth	PRV5	1016	196	5.18
Total	4	3474	748	4.58

f=frequency of questions, no= number of objectives, MoNE= Ministry of National Education, PRV=Private

According to Table 1, the numbers of questions in fourth and fifth grade science and technology textbooks presented clear differences in proportion of questions to total objective number for each grade. Based on a comparison of the two same level textbooks, it was seen that the difference between the fifth grade textbooks was 7%, while the difference between fourth grade textbooks was 20%. What is interesting about these findings is a clear difference in fourth grade textbooks. In the fifth grade textbook (MoNE5) approximately 5 questions per objective were determined, while approximately 3 questions per objective were determined in the textbook MoNE4. This result showed that there was 50% rate of difference between the numbers of questions in these two textbooks. The lowest rate of difference (11%) for the number of questions was found between PRV4 and MoNE5. Chi-square test results also showed statistically significant differences in question numbers between the two grade levels ($\chi^2 = 84.77$, $p < 0.05$).

Table 2. Distribution of Questions in Science and Technology Textbooks across Their Locations and Cognitive Levels

Domain	Beginning of unit			In-text			End of unit			Total
	Input	Processing	Output	Input	Processing	Output	Input	Processing	Output	
Physics	48	72	63	176	246	96	201	161	43	1106
Chemistry	33	46	54	175	271	126	122	85	30	942
Biology	49	88	46	194	250	102	201	108	48	1086
Astronomy	9	29	17	62	53	18	102	29	21	340
Total	139	235	180	607	820	342	626	383	142	3474

In Table 2, it is illustrated that approximately half of the questions in course textbooks are included in in-text location. Rate of the questions in end of unit is 1/3. While more than half of the question regarding astronomy is included in end of unit, number of the astronomy questions in beginning of unit is the lowest. Majority of the questions regarding biology, chemistry and physics are included in in-text location. When the distribution of the questions in beginning of unit position across the domains, it is seen that various numbers of the questions for each Costa's cognitive development level are determined. In particular, input level astronomy questions at the beginning of the units are fewer than other fields. In addition, distribution of output level questions varies across the fields. This situation causes significant differences in numbers of the questions at the beginning of the units across their levels (χ^2= 13.06, $p < 0.05$).

Table 2 represents that number of the questions rergarding different fields in intext location also varies across their level. It was found that the difference is likely related to the questions in astronomy field. While near the half of chemistry, biology and physics questions in in-text location is about process skills, near the half of astronomy questions focuse on input skills. Based on these results, it can be said that number of the questions regardin different fields differed significantly across their levels (χ^2= 14.71, p< 0.05). Finally, disrtibutions of the questions in the end of units in terms of fields and their levels were investigated and significant differences were found. As similar to in-text questions, number of the questions in the end of units differed sifnificantly for astronomy field in terms of the levels. The rates of questions regarding to physics, biology and chemistry in the end of units are similar to each other. Near the half of the questions regarding chemistry, biology and physics are related to input level, while 2/3 of the astronomy questions are at input level. These results show that numbers of the questions regarding different fields in the end of units differed significantly across Costa's cognitive development levels(χ^2= 24.20, p< 0.05). Table 3 showed distribution of questions according to their locations and cognitive levels.

Table 3. Distribution of Questions in Science and Technology Textbooks across Their Locations and Cognitive Levels

Books	Beginning of unit			In-text			End of unit			Total
	Input	Processing	Output	Input	Processing	Output	Input	Processing	Output	
MoNE4	35	72	62	103	100	32	135	92	17	648
PRV4	44	41	41	147	167	86	173	141	82	922
MoNE5	52	109	62	123	194	46	191	82	29	888
PRV5	8	13	15	234	359	178	127	68	14	1016
Total	139	235	180	607	820	342	626	383	142	3474

According to Table 3, it was seen that more than half of the questions (50.9%) were located in-text, while 33.3% of the questions were at the end of units and remaining questions were in the beginning of units. The result of chi-square test showed that locations of the questions differed significantly across cognitive levels of them (χ^2= 214.32, $p < 0.05$). Majority (42%) of the questions at the beginning of the units were processing questions while 25 % of them were input questions. For the in-text location, majority (46.4%) of the questions were processing questions while percentages of input and output questions were 34.3% and 19.3% respectively. For another location, end of the units, majority (54.4%) of the questions were input questions, whereas output questions included 12.3 % of whole questions at the end of the units. Percentage of processing questions was 33.3 %. The results presented in Table 2 were in line with the literature (Pizzini, Shepardson & Abell, 1992; Kahveci, 2010). They found that majority of the questions in the textbooks were in-text questions, which included processing questions. Results of this study also showed that 41.4 % of all of the questions in the textbooks included processing questions while 39.5 % and 19.1 % of the questions, recpectively, were input and output questions.

Being another point presented in Table 3, comparison results based on differences in the textbooks published by different authorities showed important differences in the number of questions. It was found that the highest number of questions was for PRV5, whereas the least number of questions was for MoNE4. The largest differences between textbooks were found in distribution of the questions in the beginning of the units. As another focus of the current study, question types preferred in the textbooks were presents in Table 4.

Table 4. Question Types Used in the Textbooks

Books	Question types						
	Essay	Cloze	MC	TF	Matching	Alternative	Total
MoNE4	488	57	52	43	11	2	648
PRV4	632	155	48	7	28	52	922
MoNE5	721	40	50	44	32	1	888
PRV5	817	74	25	20	35	45	1016
Total	2658	321	175	114	106	100	3474

MC: Multiple-Choice, TF: True-False

When Table 4 is considered, the distribution of questions types acroos the textbook question types was heterogeneous. Majority (76.4%) of the questions were open-ended type of questions whereas the rate of alternative evaluation questions was very low. Traditional evaluation questions comprised of 97.1% of all questions, while alternative evaluation questions including word association, project evaluation, poster evaluation, flow chart, cognitive maps and concept maps were presented at the rate of %2.9.

Based on the data presented in Table 4, first five question types were grouped under traditional evaluation questions, while the sixth question type was grouped under alternative evaluation questions. By using the chi-square technique, differences in the distribution of different questions types acroos the textbooks was investigated and it was found that the distribution of the two question types differed significantly for different textbooks (χ^2= 78.88, p< 0.05). This result showed that questions of the textbooks did not have a balanced distribution in which more alternative evaluation questions should have been included in line with the requirements of the curriculum. Therefore, the alternative evaluation, as was emphasized in the current curriculum, was not represented in the textbooks adequately.

When textbooks were considered one by one, it was seen that MoNE4 and MoNE5 had the least number of alternative evaluation questions. The most frequent type of questions in MoNE4 was essay questions (75%), while the proportion of alternative evaluation questions was 0.3%. Similarly, PRV4 included mostly essay questions (68.5%), whereas true-false questions had a rate of 0.7%. Fifth grade textbooks also presented a similar problem about lower rates of alternative evaluation questions. Majority of the questions (81.2%) in MoNE5 were essay type questions, while 0.1% of the questions were alternative evaluation questions. The other fifth grade textbook (PRV5) also included essay questions at higher rates (80.4%), while true-false questions were presented at a rate of 2.0%. These results showed that essay type questions were the most dominant type of questions and alternative evaluation questions were the least frequent questions.

Discussion and Suggestions

The results of this study demonstrated inconsistency between the number of questions per unit and the number of objectives per unit as a sign of weight of unit. It was also evident that this inconsistency was observed between the same level textbooks as well as different grade textbooks in terms of correspondence between the number of questions and objectives per unit. This result is an indication of insufficient attention to the quantity of questions in textbook writing. It might be said that differences in the number of questions in the textbooks are reflections of the author's (or authors') desire to do their own design and to apply their ideas on the textbooks. In fact, the quality of textbooks is more important than quantity. Appropriately prepared questions might function effectively and work better than two or more questions with lower quality (Haladyna, 2004). This assertion might be accepted as rationale for requirement of authors' focusing on quality. In addition to quantity analysis, quality analysis was also performed in the current study. Four different textbooks were analyzed based on Costa (1985)'s cognitive domain taxonomy in order to check the quality of questions. In general, majority of the questions were found to be located in texts position. In addition, input and processing questions were more than output questions. Examples of different cognitive level questions found in the textbooks analyzed are given in Appendix. Kahveci (2010)'s study provided conflicting results with the results of this study: Kahveci found that majority of the questions at the beginning of the units was processing questions. But this study showed that only PRV5 had a similar structure with the results of Kahveci.

Similar to the results of this study, Dávila and Talanquer (2010) compared textbooks by taking into account taxonomies of Costa and Bloom and they found that majority (4/5) of the questions corresponded to knowledge and comprehension levels, while minority (1/5) of the questions included synthesis and evaluation level questions. This result indicated that cognitive purposes of the questions should be considered in addition to their quantity.

It was found that numbers of the questions in textbooks differed significantly across the fields. In particular, distribution of the questions of astronomy field represented clearer differences than those in other fields. For this situation, basic reason is that every field is structured by using its own special purposes. For each field, different groups of field experts work, thus causing difference in numbers of the questions, as a result, there is a need to establish common frame among the experts for decreasing the differences in textbooks.

Another result of the study is that the majority of the questions (3/4) in four textbooks were essay type questions. Although essay questions are related to higher-order-thinking skills (Popham, 2003), essay questions in the textbooks are "what" questions rather than "why" and "how" questions. Predominancy of "what" questions is an indication of the insufficieny of questions for reaching objectives of the constructivist curriculum. As an important point, very low rates of alternative questions in the textbooks showed that the objectives and requirements of the curriculum were not taken into accout by textbook writers. Ministry of National Education emphasized the significance of alternative evaluation methods, but the number of questions in alternative evaluation category was very low, which leads to a controversy between the contents of the curriculum and the textbooks. Moreover, majority of the questions in textbooks focused on the measurement of output skills rather than processing skills, which was a major purpose of the curriulum (MoNE, 2005). The discrepancy between the objectives and the quality of questions, as well as the requirements of the curriculum might have been caused by resistance to change according to the new curriculum. Anderson (1998) claims that adopting alternative evaluation requires fast changes and time, after a long period of using common evaluation ways. Adoption of alternative evaluation cannot be guaranteed by only changing curriculums. Students, teachers, authorities and especially books writers should be prepared for more sustainable changes.

Despite the fact that preparation of curriculum based on the constructivist approach is a good success, lack of textbook quality to compensate requirements of the new curriculum is an important problem to effective implementation of such curriculum. In the constructivist approach, evaluation is a means of learning rather than being based on pure scores of the students. Constructivist curriculum were viewed as easily administrable by teachers and students, while their ideas were positive (Erdoğan, 2007; Şahin, 2008).

As the first step to go further in reaching special and general objectives of the science and technology education in Turkey, positive points of view on curricular attempts should be provided. One of the most important ways to provide positive ideas is to realize requirements of contemporary educational approaches including constructivism. The evaluation aspect of curriculum development is very important to conclude on the functionality of a curriulum. In other words, being an effective feeadback tool for evaluating a curriculum requires attention to problems that are described on questions of science and technology textbooks in this study. Especially, inappropriate distribution of the questions with respect to the objectives and lack of alternative type questions are important issues for both reaching the objectives of the curriculum and leading students to increase their higher-order-thinking abilities. To overcome these problems, governmental authorities should determine and publish criteria for writing textbooks in line with the curriculum. Textbooks are the basic teaching resources for majority of teachers (O'Sullivan, 2006; Radcliffe et al., 2008). Analysis of other textbooks in terms of the quality and purpose of the questions might provide a much clearer picture.

Despite the fact that this study provided fundamental insights on the problems of the quality and distribution of questions in textbooks, limited number of books being examined is the main shortcomings of this study. For more and indeepth understanding, there is an urgent need to study additional and diverse science and technology textbooks in use. Moreover, student work-texts and teacher guide-books should also be studied with similar methodologies employed in this study in order to check consistency with the curriculum objectives. Another suggestion is that verbal questions by science and technology teachers should also be analyzed by considering their relation and relevance to textbook questions and curriculum objectives.

References

AAAS. (American Association for the Advancement of Science) (2001). *Designs for science literacy.* Oxford, England: Oxford University Press.

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. *Review of Research in Education,* 42(2), 145-170.

Anderson, R.S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning,* 74, 5-16.

Armbruster, B. & Ostertag, J. (1993). Questions in elementary science and social studies textbooks. In B.K. Britton, A. Woodward & M. Binkley (Eds.), *Learning from textbooks: Theory and practice*, (pp. 69-93), Hillsdale, NY: Lawrence Erlbaum Associates, Publ.

Ayas, A., Çepni, S., Johnson, D. & Turgut, M.F. (1997). *Kimya öğretimi [Teaching Chemistry]*. YOK/World Bank National Education Devolepment Project, Bilkent, Ankara.

Berberoğlu, G. & Kalender, İ. (2005). Öğrenci başarılarının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi, [Investigation of student achievement across years, school types and regions: The SSE and PISA analyses]. *Journal of Educational Sciences and Practice,* 4(7), 21-35.

Berberoğlu, G., Kaptan, F. & Kutlu, Ö. (2002). Türkiye genelinde sekizinci sınıf öğrencilerinin fen bilgisi dersindeki üst düzey zihinsel becerilerinin incelenmesi. [Investigation of higher order cognitive abilities of eighth graders on science content: Turkish case]. *Fifth National Science and Math Education Congress,* Middle East Technical University, September 16-18, Ankara, Turkey.

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Brill, G. & Yarden, A. (2003). Learning biology through research papers: A stimulus for question-asking by high-school students. *Cell Biology Education,* 1(2), 266-274.

Colburn, A. (2000). Constructivism: Science education's grand unifying theory. *The Clearing House*, *74*(1), 9-12.

Costa, A.L. (1985). Teacher behaviors that enable student thinking. In A.L. Costa (Ed.), *Developing minds: A resource book for teaching thinking,* (pp. 125-137), Alexandria, VA: Association for Supervision and Curriculum Development.

Dávila, K. & Talanquer, V. (2010). Classifying end-of-chapter questions and problems for selected general chemistry textbooks used in the United States. *Journal of Chemical Education,* 87(1), 97-101.

Erdoğan, M. (2007). An analysis of a newly developed fourth and fifth grade science and technology course curriculum: A qualitative study. *Journal of Turkish Educational Sciences,* 5(2), 221-254.

Gall, M.D. (1970). The use of questions in teaching. *Review of Educational Research,* 40(5), 707-721.

Gelbal, S. & Kelecioglu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar [Teachers' proficiency perceptions of about the measurement and evaluation techniques and the problems they confront]. *Hacettepe University, Journal of Education Faculty,* 33, 135-145.

Gür, B.S., Çelik, Z. & Özoğlu, M. (2012): Policy options for Turkey: A critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy,* 27(1), 1-21.

Haladyna, T.M. (1992). The Effectiveness of several multiple-choice formats. *Applied Measurement in Education,* 5(1), 73-88.

Haladyna, T.M. (2004). *Developing and validating multiple-choice test items*. (3^{th} ed.), New Jersey: Lawrence Erlbaum Associates, Publishers.

Jupp, V. (Ed.). (2006). *The Sage dictionary of social research methods*. London: SAGE Publications Ltd.

Kahveci, A. (2010). Quantitative analysis of science and chemistry textbooks for indicators of reform: A complementary perspective. *International Journal of Science Education,* 32(11), 1495-1519.

Krippendorff, K. (2003) *Content analysis: An introduction to its methodology.* (2^{th} ed.), Beverly Hills, CA: Sage Publication, Inc.

Meyer, L.A., Crummey, L. & Greer, E.A. (1988). Elementary science textbooks: Their contents, text characteristics, and comprehensibility. *Journal of Research in Science Teaching,* 25, 435-463.

MoNE (Ministry of National Education). (2005). *İlköğretim fen ve teknoloji dersi (4. ve 5. sınıflar) öğretim programı.* Retrieved February 19, 2013, from' http://ttkb.meb.gov.tr. index1024.htm. (in Turkish)

MoNE (Ministry of National Education). (2007). *ÖBBS raporu: İlköğretim öğrencilerinin başarılarının belirlenmesi Türkçe raporu*. Turkish Ministry of Education Education Research and Development Department, Ankara. Retrieved February 02, 2013, from' http://earged.meb.gov.tr. (in Turkish)

Newport, J.F. (1965). An evaluation of selected series of elementary school science textbooks. *Science Education,* 49(5), 479-484.

NRC (National Research Council). (1996). National Science Education Standards. Washington, DC: National Academy Press.

Oakes, J. & Saunders, M. (2004). Education's most basic tools: Access to textbooks and instructional materials in California's public schools. *Teachers College Record,* 106(10), 1967-1988.

Ogan-Bekiroglu, F. (2007). To what degree do the currently used physics textbooks meet the expectations? *Journal of Science Teacher Education,* 18(4), 599-628.

O'Sullivan, M. (2006). Lesson observation and quality in primary education as contextual teaching and learning processes. *International Journal of Educational Development,* 26, 246-260.

Pizzini, E.L., Shepardson, D.P. & Abell, S.K. (1992). The questioning level of select middle school science textbooks. *School Science and Mathematics,* 92, 74-79.

Popham, W.J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Radcliffe, R., Caverly, D., Hand, J. & Franke, D. (2008). Improving reading in a middle school science classroom. *Journal of Adolescent and Adult Literacy,* 51(5), 398-408.

Rourke, L. & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development,* 52(1), 5-18.

Şahin, İ. (2008). The evaluation of new science and technology curriculum for grades 4 to 5. *Milli Eğitim Dergisi (The Journal of National Education), 177,* 181-207.

Shepardson, D.P. & Pizzini, E.L. (1991). Questioning levels of junior high school science textbooks and their implications for learning textual information. *Science Education,* 75, 673-682.

Stake, R.E. & Easley, J.A. (1978). *Case studies in science education.* Center for Instructional Research and Curriculum Evaluation, Urbana: University of Illinois.

Stern, L. & Ahlgren, A. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching,* 39(9), 889-910.

Tavşancıl, E. & Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri [Content analysis of verbal, written and other materials and application examples]*. Istanbul: Epsilon Publishing.

Wang, H.A. (1998). *Science textbook studies reanalysis: Teachers "friendly" content analysis methods.* Paper presented the Annual Meeting of NARST, 71st, April 19-22, San Diego, CA, (ERIC Document Reproduction Service No. ED423142).

Yager, R.E. (1996). *Science/technology/society as a reform in science education.* New York: State University Press.

Yılmaz, A., Seçken, N. & Morgil, İ. (1998). Lise 11. sınıf, kimya 3 ders kitaplarının kimya eğitimine uygunluklarının araştırılması. *Hacettepe University, Journal of Education*, 14, 73-83. (in Turkish)

Appendix. Question examples according to Costa's cognitive domain taxonomy from Science and Technology Textbooks

Books	MoNE4	PRV4	MoNE5	PRV5
Input	What is the factor affecting the shape of earth's crust in an earthquake? (p. 121)	List the functions of muscles in our body. (p. 17)	Which of the following is not one of our energy resources? a) Lamp b) Coal c) Gasoline d) Sun (p. 103)	What is to change from liquid to solid called? (p. 73)
Processing	You buy rice with kilo and caffee with gram. What is this difference? (p. 69)	Compare plants and animals according to their nutrition. (p. 165)	Can we examine dove and bat in the same group? (p. 167)	How does brightness of the bulbs change as the number of them increase in series circuit? (p. 116)
Output	Explore the reason of mildew for the food left on the counter. (p. 217)	Explore the reason for the liquid oil poured minto mold becoming solid. (p. 77)	What is the best ways of getting rid of hazardous substances? (p. 41)	How do you explain that one side the Earth being dark constantly? (p. 227)