Cognitive level analysis of science item tests on secondary school assessment

I Prihastuti1*, A Widodo1,2 and Liliasari1,3

1Program Studi Pendidikan IPA, Sekolah Pascasarjana, Universitas Pendidikan Indonesia, Jl. Setiabudhi No. 229, Bandung, 40154, Indonesia
2Departemen Pendidikan Biologi, Universitas Pendidikan Indonesia, Jl. Setiabudhi No. 229, Bandung, 40154, Indonesia
3Departemen Pendidikan Kimia, Universitas Pendidikan Indonesia, Jl. Setiabudhi No. 229, Bandung, 40154, Indonesia

*Corresponding author’s email: arifinindrie@gmail.com

Abstract. This case study aims to obtain an overview of cognitive levels in science items tests used in the secondary school assessments in Banten Province. Cognitive level analysis was carried out on 3 types of science assessment documents (1) national examination (UN) from year 2017 and 2018; (2) national standardized school examination (USBN) used in eight districts from year 2018, and (3) summative tests taken from 8 secondary schools in each district of Banten Province. Determination of cognitive level was using Bloom taxonomy revision. In addition, data collection was conducted through questionnaires to capture information about teacher professional development on designing science items test. The questionnaire was responded by 30 science teachers from representatives of 4 districts in Banten province. Interview was conducted for 5 selected teachers to get deeper information about professional development programs. The results showed that there were no higher-level questions which made by teachers in USBN and summative test documents. The teacher’s professional development programs related to assessment and test item development were still lacked of direct practices especially for developing higher level questions. Intensive teacher professional development programs are really needed to improve teachers’ skills in developing higher level items test.

1. Introduction

One of the 21st century essential skills are learning and innovation skills. These skills include: critical thinking and problem solving, creativity and innovation, communication and collaboration [1]. Critical, creative thinking and problem solving skills should be represented in primary and secondary curricula [2]. The curriculum in Canada for example, higher level thinking is the most fundamental goal in all subjects [3]. Higher level terminology was identified in curriculum documents in four Australian states [4]. As many as 86.7% of learning objectives in the science curriculum tend to be at cognitive level of understanding and application and some learning objectives in the science curriculum in Korea are included in the level of creation [5]. Some references to science learning standards, such as National Science Education Standard (NSES) or Next Generation Science Standard (NGSS), were emphasizing on students ability to solve problems in various situations. These standards also illustrate that higher level thinking is an important aspect in science learning [6]. In Indonesia, the
2013 curriculum has adopted Bloom's taxonomy revision by Anderson which starts from the level of knowing, understanding, applying, analyzing, evaluating and creating. Higher level thinking demands on students are reflected in the Graduate Competency Standard (SKL) document in 2013 Curriculum [7].

The dominance of assessment practices that focus on lower level thinking has been a concern among researchers [8]. There were many teachers who do not understand how to teach and assess higher level of thinking, resulting on students low-level thinking skills [3]. This was indicated by students who were still having difficulty on analyzing information, tend to accept what information is obtained, passive in asking questions or answering questions from problems raised by teachers, and passively on expressing ideas to solve problems [9]. The 2015 TIMSS results showed that Indonesian students only master a routine, simple computing, and knowledge about day-to-day facts. Indonesian students need to strengthen their ability to integrate information, draw conclusions, and generalize their knowledge to solve the problems [10]. Related to evaluation of students learning, the Ministry of Education and Culture (MoEC) has been conducted the national examination (UN) and the national standardized school examination (USBN) at the end of school year. UN is a tool on mapping and assuring the quality of education which administered by MoEC [11]. Meanwhile, USBN is an evaluation to measure student competency achievement and as a determinant of students graduation which administered by the unit of education [11]. As a benchmark for SKL achievement, these two forms of assessment should reflect questions that measure higher level thinking skills which demanded in SKL. In order to answer those type of questions, students should be accustomed to face this type of questions in their daily learning at school. One of the efforts to generate students with higher level thinking skills is learning instruction should be able to train those skills through teachers made questions. However, many studies have showed that teachers at all levels of education tended to ask mostly questions requiring recall and rarely ask questions requiring analysis [12-14]. Therefore, this research will be conducted to reveal in what extent teachers questioning in their assessment documents (USBN and summative test) have fulfil the higher level thinking demands. In addition, questions in UN document will be analysed as well to get information about the proportion of higher cognitive level questions in the last two years. The discussion of this study will be focused on the cognitive level of question based on Bloom's taxonomy revision [15]. For supporting data, teachers’ professional development activities related to assessment and designing item test will be collected. The research questions that will be discussed in this study were: (1) how does the cognitive level of science item test on secondary school assessments; (2) How does professional development programs support teachers’ skill in designing higher level items test.

2. Methods
This research is a case study in Banten Province with document analysis as the main data collection. Cognitive level analysis was carried out on 3 types of science assessment documents, (1) national examination (UN) from year 2017 and 2018; (2) national standardized school examination (USBN) item tests used in eight districts from year 2018, and (3) summative item tests, taken from 8 secondary schools in each district of Banten Province. Determination of cognitive level was using Bloom taxonomy revision [12]. In addition, data collection was conducted through questionnaires to capture information about teachers’ professional development on designing science items test. Questionnaires were disseminated through online media to facilitate data collection. The questionnaire was responded by 30 science teachers from representatives of 4 districts in Banten province, namely Lebak, Serang, Cilegon and Tangerang. Interview was conducted for 5 selected teachers to dig up deeper about the answers from questionnaire. The teachers were asked to explain more deeply about the obstacles in designing higher level questions and to what extent the professional development activity can support their assessment tasks. All data collection was analysed descriptively.
3. Results and Discussion

3.1. Cognitive levels in assessment documents

3.1.1. UN documents. UN documents from year 2017 and 2018 consist of 40 MCQs respectively. Based on the results of the analysis, there are four cognitive levels contained in those questions, the knowledge (C1), understanding (C2), applying (C3) and analysis (C4) level. The results of the analysis are shown in figure 1 below:

![Figure 1. Percentage of cognitive level in UN document](image)

The highest percentage of cognitive levels is level C2 (understanding) as much as 58%, both in 2017 and 2018 documents. The next percentage is cognitive level C3 (Applying), where the percentage in 2017 is 25% and increased to 28% in 2018. The third percentage is level C4 or analysing, where the number increases to 15% compared to the previous year which was only 10%. There were still level C1 (remembering) questions in 2017 document, but this type of question doesn’t appear again in 2018. These results indicate that the percentage of cognitive levels specifically for the HOTS category (level C3 and C4), has increased over the past two years. Meanwhile, lower level questions (level C1) were decreasing compared to the previous year. Similar to Dongoran study on UN science document from year 2012 to 2014, especially for Biology content, the questions only reached cognitive level C4, meanwhile level C5(evaluating) and C6 (creating) are not found in these documents [16].

The development of assessment tools are inseparable from the competency standard for science curriculum which set by Ministry of Education and Culture [7]. Competency standards (SK) and basic competencies (KD) are minimal competencies that students must achieve during their school grade level. However, teachers can develop learning instructions and assessments more than competency described in KD. Each competency using verb that describing cognitive level achievement. Based on the verbs listed in the basic competencies, there are six types of verbs namely: analysing (17 KD), explaining (6 KD), applying (5 KD), connecting (3 KD), identifying (2KD) and classifying (1 KD). Verbs that belong to the HOTS category are analysing, applying and connecting [15]. KD with this verb has the highest number compared to the number of KD that has the LOTS verb. This shows that the Indonesian government has tried to drive teachers in developing high-level thinking skills on learning and assessment. The increasing number of questions with analysing category in two previous years showed that teachers must prepare their students to face these kind of questions. It is very possible that in the following years, there will be more questions that measure higher level thinking skills considering that these skills are curriculum demands. The questions at cognitive level of evaluating (C5) and creating (C6) have not yet appeared in the basic competencies listed in science curriculum. This might cause the emergence of questions in these categories in the UN document.

3.1.2. USBN documents. The composition of cognitive levels in USBN documents are shown in Figure 2 below. Questions with cognitive level C2 has the highest composition as much as 50%, followed by C3 level of 26%, C1 level of 18% and the C4 level of 6%. The number of questions for each level varies in each district, except for questions with level C4. The percentage of level C1 questions ranged from 5-29%, C2 ranged from 45-61%, C3 ranged from 21-29% and level C4 were 6% for each district.
Since 2017 the government has enacted a policy to implement USBN as an effort to improve the quality of school examinations. Science subjects are one of those tested on USBN since 2018. Differently from UN, majority (75-80%) of USBN questions were made by teachers who were members of the MGMP forum, the rest (20-25%) were made by the government through MoEC as anchor items. One of the benefits of implementing USBN is to increase teacher and MGMP capacity in the designing and organizing item test. The policy is expected to increase teachers’ activity in making good quality item test which will be used for their daily assessment. Based on the analysis, majority of the questions (level C1-C3) were made by teachers. The 6% of level C4 questions were part of the 20-25% of questions made by MoEC. It was indicated that none of higher level question were made by teachers.

3.1.3. Summative test document. The results of cognitive level analysis on the summative test document revealed that level C3 questions had the highest proportion of 56%, followed by level C1 and C2 questions with the same proportion of 22% (Figure 3). None of the questions were belong to levels C4, C5 and C6.

![Figure 2. Proportions of cognitive level in USBN documents](image1)

![Figure 3. Proportion of cognitive levels on summative assessment](image2)
As explained above, the number of KDs with analyzing verbs has the highest number compared to lower thinking categorized verbs. Teachers seems too busy with the day-to-day role of supplying information to students and did not take the time to reflect on higher order processes in science, such as how to analyze evidence in decision making [17]. Books for students and teachers which published by government to be used in schools, have led teachers to carry out instruction that training students to have critical and creative thinking skills. The development of these skills also requires appropriate assessment in order to provide suitable description of student achievement and recommendation for education improvement [18]. However, the development of assessments that can measure critical thinking skills cannot be done by the teacher. One reason is that the teacher does not have a guideline to develop higher-level thinking item test [19]. Teachers were usually taken items test from various textbooks or item collection books. It will lead to misalignment between curriculum outcomes and assessment. The alignment of outcomes and assessments has “the potential to have a positive impact on all students’ learning and achievement” [20]. Based on the results of the analysis on the UN question, higher level thinking questions that were inserted in the UN questions continued to increase. Therefore, the average test scores carried out by schools should not differ from the average score of the National Examination. Consequently, higher level thinking instruction and assessment should be inserted in student’s daily classroom activities [21]. The teacher's skills in developing high-level thinking questions in science subjects really need to be improved, in order to practice the students with higher level thinking through these questions.

3.2. Teachers professional development activities

Based on the results of the questionnaire, most of the teachers (73%) were involved in professional Development (PD) programs related to assessment. Teachers were participating in this activity 1-2 times on average in the last two years. The analysis revealed that 57% of teachers have received information about HOTS. 67% teachers claimed that they have inserted HOTS questions in their test items, but it was not reflected in their summative test documents. They were not trained in depth about how to design HOTS questions. Most of the PD programs related to assessment were MGMP program with the duration ranging from 3-4 hours. The other activities were obtained through the 2013 curriculum training program with 6 hours duration (only for assessment). Those programs were lack of time and practices to fulfil teachers’ skills in developing assessment. School programs related to assessment PD were occasionally happen when semester test has nearly come. Professional development activities were only learn about theory but very lacking in developing practical skills in the classroom. Training or workshops are rarely followed up with coaching or mentoring. Consequently, teachers often feel confused in applying the new knowledge that they get at the training. As a result, the teachers returned to their old style of instruction and assessment.

A more intensive programs in guiding teachers to develop HOTS questions was a mentoring program for schools model administered by Educational Quality Assurance Institution (EQAI) in Banten province. However, this program was mentored by one instructor and applied for all teacher in one school with 3 times of mentoring (3-4 hours each time). It was still less intensive and did not apply for all schools in Banten Province. There were many schools who need for this assistance from EQAI to improve their teachers quality. According to Young's, the one shot workshops or trainings were unable to change the way teachers taught to improve their students' skills and achievements [22]. The study by Corcoran found that teachers who participated in professional development activities for ≥80 hours, significantly showed the use of teaching practices according to what they had learned compared to teachers who took <80 hours of activity [23]. From the interview, teachers felt more comfortable with mentoring programs than delivering information through workshops or trainings. Trueusdale revealed that teachers who were accompanied by coaches or assistants can apply the knowledge in their classrooms, while teachers who only get knowledge from workshops did not practice it in their classroom [24].

Based on the results discussed above, suggested recommendation for further improvements are: (1) MGMP activities in designing HOTS questions should be a routine and intensive programs, not only
carried out before USBN but the product of this program can be used for their daily assessments; (2) EQAI can develop their programs by facilitating MGMP through a mentoring program especially in designing HOTS instruction and assessment; (3) This research could be used to review and refine teachers assessment documents to be align with KD achievement

4. Conclusion
Science item test with higher cognitive levels in UN documents were increasing in the past two years. Questions made by teachers only reach level C3 in USBN document and summative test. There were no higher level questions which made by teachers. Misalignment of cognitive level was occurred between KD and teachers questions in summative test documents especially for KD with analyzing verbs. The teachers professional development programs related to assessment and test item development were still lack of direct practices especially in developing HOTS items. It is necessary to develop teachers’ professional development program to improve teachers' skills in developing higher level questions by using guidance and coaching strategy.

5. References
[1] Trilling B and Fadel C 2009 21st Century: Learning for Life in Our Times (San Fransisco: John Wiley & Sons)
[2] Fisher R 2005 Teaching children to think (England Nelson: Thones Ltd)
[3] FitzPatrick B Schulz H 2015 Canadian Journal of Sci. Math and Tech. Education 15 136
[4] Fensham P P and Bellochi A 2013 Thinking Skills and Creativity 10 250
[5] Lee Y J, Kim M and Yoon H G 2015 International Journal of Science Education 37 2193
[6] Zohar A 2004 Higher Order Thinking in Science Classroom: Students’ Learning and Teachers’ Professional Development (New York: Springer)
[7] Kemdikbud 2013 Standar Kompetensi Lulusan (Jakarta: Kemendikbud)
[8] Chen X, Jiang M, Cai L, Liang L L, Du J and Zhou Y 2017 Contemporary Trends and Issues in Science Education 45 235
[9] Himah F and Sudarti S 2016 Jurnal Pemelajaran Fisika 5 89
[10] Rahmawati 2016 TIMSS 2015 Report Analysis (Jakarta: Puspendik)
[11] BSNP 2018 Prosedur Operasional Standar Penyelenggaraan Ujian Sekolah Berstandar Nasional (POS USBN)2018 (Jakarta: Kemdikbud)
[12] Inamullah H M 2011 Asian Social Science 9 149
[13] Iskandar D Senam S 2015 Innovation of Science Education Journal 1 65
[14] Hill J D Flynn K 2008 Journal of Staff Development 29 46
[15] Anderson L W, Krathwohl D R 2001 A taxonomy for learning, teaching and assessing: A revision of Bloom’s taxonomy of educational objectives (New York: Longman)
[16] Dongoran J 2015 Analisis Butir Soal Ujian Nasional SMP dan Soal PISA pada Konten Pengetahuan Biologi Berdasarkan Taksonomi Bloom Dua Dimensi (UNIMED)
[17] Levinson R 2006 The Curriculum Journal 17 247
[18] Abosalem Y 2016 International Journal of Secondary Education 4 1
[19] Jusuf R, Sopandi W, Ratnavulan A, Syaefudin S 2018. Jurnal Visipena 9 1115
[20] Roach AT, Niebling BC and Kurz A 2008 Psychology in the Schools 45 158
[21] Kemdikbud 2017 Modul Penyusunan Soal Higher Order Thinking Skills (Jakarta : Kemendikbud)
[22] Youngs P 2013 Using teacher evaluation reform and professional development to support Common Core assessments (Washington DC: Center for American Progress)
[23] Corcoran T, McVay S and Riordan K 2003. Getting it right: The MISE approach to professional development (Philadelphia: Consortium for Policy Research in Education)
[24] Truesdale W T 2003 The implementation of peer coaching on the transferability of staff development to classroom practice in two selected Chicago public elementary schools (Univ of Chicago)
Acknowledgments
Authors wishing to acknowledge teachers who have participated in the surveys, colleagues and lecturers who always give their encouragements from the beginning until the end of the research and also EQAI in Banten Province for the technical support that make this research can be done properly.