

STUDY OF ITEMS ABOUT DAILY ASSESSMENT MADE BY TEACHERS OF PLANT TISSUE STRUCTURE AND FUNCTION MATERIAL GRADE XI SMA SCHOOL YEAR 2020/2021

Telaah Butir Soal Penilaian Harian Buatan Guru Materi Struktur dan Fungsi Jaringan Tumbuhan Kelas XI SMA Tahun Ajaran 2020/2021

Faradillah Asri

Biology Education, Faculty of Mathematics and Natural Science

E-mail : faradillah.17030204083@mhs.unesa.ac.id

Raharjo

Biology Education, Faculty of Mathematics and Natural Science

E-mail : raharjo@unesa.ac.id

Abstract

This research is motivated by the low percentage of daily assessment of plant tissue structure and function material of 37.8%. The daily assessment problem that has been tested to students has never been analysis, therefore researchers aim to describe the quality of the problem, improve the problem and obtain better quality questions in the daily assessment made by teachers of plant tissue structure and function materials at SMAN 20 Surabaya school year 2020/2021. This research is descriptive research with qualitative and quantitative methods. Data collection techniques are carried out through interview techniques, studies, and tests using the instrument sheet daily assessment made by teachers 25 multiple choice questions and 4 essays, problem improvement sheets, latticework made by teachers, and student answer sheets. The subjects of this study were 30 students from grade XI IPA 3 SMAN 20 Surabaya. The results showed the reliability value of the teacher's problem was 0.64, the difficulty level showed 2 difficult questions (8%), 9 medium questions (35%), 7 easy questions (28%) and 7 very easy questions (28%). the differentiating index shows 8 very good questions (32%), 6 good questions (24%), 4 questions are enough (16%) ; distractors show 3 very good questions (12%), 2 good questions (8%), 6 questions are enough (24%), 5 questions are bad (20%), and 9 questions are very bad (36%). Daily assessment questions of function material and plant tissue structure made by teachers can be concluded to have low reliability, low difficulty level, good difference index, and bad distractor. Improvement item can be concluded to have high reliability, medium difficulty level, highly good difference index, highly good distractor.

Keywords: Study of item question , Daily assessment, Plant tissue structure and function

Abstrak

Penelitian ini dilatarbelakangi oleh rendahnya persentase penilaian harian struktur jaringan tanaman dan fungsi bahan sebesar 37,8%. Soal asesmen sehari-hari yang diujikan kepada siswa belum pernah dianalisis, oleh karena itu peneliti bertujuan untuk mendeskripsikan kualitas soal, memperbaiki masalah dan mendapatkan soal yang lebih berkualitas dalam asesmen harian yang dilakukan oleh guru materi struktur dan fungsi jaringan tumbuhan di SMAN. 20 Surabaya tahun pelajaran 2020/2021. Penelitian ini merupakan penelitian deskriptif dengan metode kualitatif dan kuantitatif. Teknik pengumpulan data dilakukan melalui teknik wawancara, pembelajaran, dan tes dengan menggunakan instrumen lembar penilaian harian yang dibuat guru sebanyak 25 soal pilihan ganda dan 4 karangan, lembar perbaikan masalah, kisi-kisi yang dibuat oleh guru, dan lembar jawaban siswa. Subjek penelitian ini berjumlah 30 siswa kelas XI IPA 3 SMAN 20 Surabaya, Hasil penelitian menunjukkan nilai reliabilitas soal guru sebesar 0,64, tingkat kesulitan menunjukkan 2 soal sulit (8%), 9 soal sedang (35%), 7 pertanyaan mudah (28%) dan 7 pertanyaan sangat mudah (28%). indeks pembeda menunjukkan 8 pertanyaan sangat baik (32%), 6 pertanyaan baik (24%), 4 pertanyaan cukup (16%); distraktor menunjukkan 3 pertanyaan sangat baik (12%), 2 pertanyaan baik (8%), 6 pertanyaan cukup (24%), 5 pertanyaan buruk (20%), dan 9 pertanyaan sangat buruk (36%). Soal penilaian harian materi fungsi dan struktur jaringan tanaman yang dibuat oleh guru dapat disimpulkan memiliki reliabilitas rendah, tingkat kesulitan rendah, indeks selisih baik, dan distraktor buruk. Item perbaikan dapat disimpulkan memiliki reliabilitas tinggi, tingkat kesulitan sedang, indeks selisih sangat baik, distraktor sangat baik.

Kata Kunci: Telaah, Penilaian Harian, Struktur dan fungsi jaringan tumbuhan

INTRODUCTION

Education is one of the important sectors in creating human beings who have intellectual, social care, and moral awareness. These three aspects need to be owned, if only the intellectual aspects are owned, it will create people who do not have concern for the surrounding environmental conditions. According to their age level, each individual pursues an education with the help of several elements in it. One element that has an important role is the teacher. Teachers make great efforts in creating a quality teaching and learning process with various learning tools and media. In addition, teachers also have several tasks, one of which is to evaluate learning.

Evaluation is a process that is carried out continuously to find out the success rate, the level of atability of learning objectives, and the constraints that exist during the learning process based on a certain criteria. When a learning process is successful it can be improved, and vice versa when encountering obstacles in learning, it can be sought solutions (Haryanto, 2020). Evaluation has 4 stages including collecting information, processing information, shaping considerations and making decisions. In a narrower scope than evaluation there is a process called assessment.

Assessment is an activity to collect, obtain data and all information about the teaching and learning process, assessment is not only in the form of grades and numbers but also the process of students in obtaining it. Such observations made by teachers to assess student behavior are also called assessment (Istika *et al.*, 2019). In the narrower scope of the assessment there is a process called tests (Novianti & Siburian, 2018).

A test is a series of assignments or questions created by the teacher and needs to be answered by the student, having a function as a measuring instrument that can determine the student's understanding of a material then compared to a predetermined grade standard. In contrast to the assessment the test only includes numbers and grades and does not look at the process (Wijayanto *et al.*, 2016). There are 2 types of tests, namely standardized tests and un standardized tests (Sudjiono, 2012). Standardized test is a test that has been done analysis including validity and reliability, such as the National Examination and the state college entrance selection test. On the contrary, un standardized tests are those that have not been analyzed both validity and reliability (Supiyansyah *et al.*, 2017). Teacher-made test is a test has not been standardized so to determine the quality needs to be done item analysis of the question both qualitatively and quantitatively. Qualitative analysis is

reviewed based on material, construction and language aspects. And quantitative analysis is obtained through the results of validity, reliability, differentiating index, difficulty level, and distractor. Validity is the accuracy of the question item in carrying out its measuring function, reliability is the determination of the item of question (Sukmela & Nofrion, 2018).

Teachers conduct several tests in 1 semester, one of which is Daily Assessment with implementation time after one material is completed. Daily assessment is important to be done as the basis for obtaining the next material. If students do not understand only 1 material, students will have difficulty in doing the test which is a collection of several materials such as midterm assessment and final semester assessment.

SMAN 20 Surabaya has a KKM (Minimum Completedness Criteria) score of 75. Daily Assessment of students of plant tissue structure and function material in grade XI IPA 3 of 62.2% below KKM, so that only 37.8% have achieved the value of completedness. The class average of 68.

In addition, the category of daily assessment made by teachers is dominant in the category C1-C3. So it is also necessary to improve the details of the questions into hots categories to obtain better quality learning results. There are eleven indicators about HOTS including focusing on questions, considering trustworthy, considering observation reports, analyzing arguments, determining conclusions, comparing conclusions, assessing, defining assumptions, defining concepts, describing (Yuniar *et al.*, 2015).

After conducting an interview with a biology teacher at SMAN 20 Surabaya there is 1 material that has never been analysis about the material structure and function of plant tissue. KD structure and function of plant tissue is 3.2 Analysing the relationship between cell structure in plant tissue and organ function in plants. Material structure and function of plant tissue is the basis for plant growth and development material in class XII. Therefore it is necessary to understand the material structure and function of plant tissues. In addition, the material that has never been analysis causes the unknown quality of each question and indirectly also affects the unknown success of the learning process.

Research related to the analysis of the problem points has been conducted by Prasetyo (2013) with the results: difficulty rate of 0% is very difficult, 17.5% difficult, 40% medium, 37.5% easy and 5% very easy; different power points about 7.5% very good, 50% good, 22.2% enough, 15% bad and 5% very bad; as much as 80% of the problem item outwits work, and has a high reliability

index of 0.727. So that the quality of daily assessment questions made by teachers can be known. Furthermore, decisions can be taken to be further improved or improvements need to be made.

Based on these descriptions, researchers are interested in conducting research aimed at research to describe the quality of the daily assessment questions made by teachers and explain the quality of teacher-made questions that have been improved in terms of content, language, and structure of plant tissue structure and function at SMAN 20 Surabaya school year 2020/2021

METHODS

This research is descriptive research with qualitative and quantitative methods because it is carried out by observing the processes and problems that exist in the community. The research was conducted in November - January 2020 at SMAN 20 Surabaya. The target of this research is about the Daily Assessment made by teachers as many as 25 multiple choice questions and 4 essay questions, answer keys, test answer sheets, and teacher-made grids. The study subjects were 60 participants in class XI IPA 3 at SMAN 20 Surabaya. Data collection techniques in the form of interviews, studies, and tests.

Qualitative data analysis techniques are carried out by 1 material expert lecturer and 1 expert lecturer of education. The review is conducted based on material, construction, and language aspects through validated instrument sheets and questions. After the instrument is validated, the score of each question item contained in the instrument using guttman measurement scale

Tabel 1. Guttman Measurement Scale

| Criteria | Score |
|----------|-------|
| Agree | 1 |
| Disagree | 0 |

Resources : Riduwan (2013)

Formula calculation of validity feasibility results from both expert lecturers:

$$\text{feasibility (\%)} = \frac{\text{obtained score}}{\text{maksimal score}} \times 100\%$$

Here is the table of interpretation of validity score:

Tabel 2. The Table of Interpretation of Validity Score

| Feasibility (%) | Criteria |
|-----------------|----------------|
| ≤ 25 - 38 | Highly invalid |
| 39 - 52 | Invalid |
| 53 - 69 | Valid enough |
| 70 - 83 | Valid |
| 84 - 100 | Highly valid |

Resources : Arifin (2011)

Quantitative data analysis techniques are carried out using Anates V4. Aspects reviewed include reliability, different index, difficulty level, and distractor. These four aspects are determined by inputting the overall value of participants and then clicking "process all automatically" then automatically will appear these four aspects. Reliability is the determination of test results if given to the same participant but at different times. Here are the criteria for test reliability:

Tabel 3. The Criteria for Test Reliability:

| Correlation Coefficient KR-20 | Reliability Criteria |
|-------------------------------|----------------------|
| > 0,70 | Reliable |
| < 0,70 | Not reliable |

Resources : Rahayu dan Djazari (2016)

The difficulty level of the question item is determined from the percentage of correct participants' answers divided by many students who took the test (Fitriatun & Sukanti, 2013). Here are the difficulty index criteria. The proportion of good difficulty is 25% difficult questions, 50% medium problems, and 25% easy questions (Zainal Arifin, 2011).

Tabel 4. The Criteria of The Difficulty Index

| Proportion | The difficulty index |
|-------------|----------------------|
| 0,00 - 0,30 | Difficult |
| 0,31 - 0,70 | Medium |
| 0,71 - 0,90 | Easy |
| 0,91 - 1,00 | Very easy |

Resources : Arifin (2011)

The different index is the ability of the problem in distinguishing students who have understood the material with students who do not understand the material. The index differs between - 1.00 to 1.00. The higher the index, the better the problem (Zaenal Arifin, 2017). Here are the different index criteria:

Tabel 5. The Different Index Criteria

| Different Index | Reliability Criteria |
|------------------|----------------------|
| DP ≤ 0,00 | Very bad |
| 0,00 < DP ≤ 0,20 | Bad |
| 0,20 < DP ≤ 0,40 | Enough |
| 0,40 < DP ≤ 0,70 | Good |
| 0,70 < DP ≤ 1,00 | Highly good |

Resources : Haryanto (2020)

Distractor can be determined by calculating the number of participants who choose each option provided

(Fitriatun & Sukanti, 2013 distraktor is said to work well if at least selected by 5% of participants. Here are the criteria for aspects:

Tabel 6. Distractor Criteria

| Malfunctioning distractor | Criteria |
|---------------------------|-------------|
| 0 | Highly good |
| 1 | good |
| 2 | enough |
| 3 | bad |
| 4 | Very bad |

Resources : Haryanto (2020)

RESULTS AND DISCUSSIONS

Point Problem daily assessment made by teachers material structure and function of plant tissue consists of 25 questions of multiple choice and 4 essays with 4 IPK (Competency Achievement Indicators) in it. Here's the IPK about the teacher's work.

Table 7. Indicators of Competency Achievement on Daily Assessment made by teachers

| Nomor | Competency Achievement Indicators (IPK) |
|-------|--|
| 3.2.1 | Describe the types of tissues in plants |
| 3.2.2 | Analyzing cell structures in various plant tissues |
| 3.2.3 | Correlate the structure and function of tissues in plants |
| 3.2.4 | Comparing tissue structures in monocotile and dicotil plant organs |

Based on the daily assessment grid of plant tissue structure and function materials made by teachers there are 4 IPK with 29 indicators of the problem contained in each problem number. IPK 3.2.1 is found in the question number 1,2,3,4,5,6,7,8,9,10,11, 1 essay and 3 essay. IPK 3.2.2 is found in the questions number 12,15,16,18, 19,20,21,25. IPK 3.2.3 is found in question number 13,14,17,22 and 2 essays. IPK number 3.2.4 is found in the numbers 23,24 and 4 essays

In IPK 3.2.1 there are 2 questions that deviate from IPK. While in other indicators there is no problem that deviates from IPK. The distorted problem is the question number 3 and number 9. Due to the discrepancy between the IPK and the item of the question made.

Table 8. Cognitive level grouping of Daily Assessments made by teachers

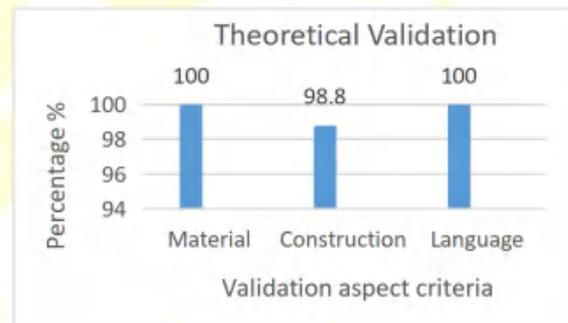
| IPK | ΣQuestion | Cognitive Level | | | | | |
|-------|-----------|-----------------|----|----|----|----|----|
| | | C1 | C2 | C3 | C4 | C5 | C6 |
| 3.2.1 | 10 | - | 1 | 5 | 4 | - | - |

| | | | | | | | |
|--------------|----|----|----|----|----|---|---|
| 3.2.1 | 11 | 5 | 5 | 2 | - | - | - |
| 3.2.2 | 8 | - | 2 | 4 | 3 | - | - |
| 3.2.3 | 4 | - | 2 | 4 | - | - | - |
| 3.2.4 | 2 | - | 2 | - | - | - | - |
| Σ | 25 | 5 | 11 | 10 | 3 | - | - |
| Percentage % | | 18 | 37 | 34 | 10 | - | - |

Based on the analysis that has been done on the cognitive level of the teacher's daily assessment, most of the questions are still dominated by the cognitive categories C2 and C3. C2 cognitive rate is 40%, C3 cognitive level is 36%, C1 cognitive rate is 16%. And the least category C4 rate is 8%. So the item questions of HOTS categories contained in the C4, C5, and C6 levels is only 8% of the total test. A total of 92% of teachers' Daily Assessments fall into the category of LOTS. This indicates that it is necessary to make improvements to the problem.

After the improvement of the question obtained a valid question with a validity value of 99.7% or very valid. Validity results are carried out by 2 expert lecturers based on material, construction and language aspects. Can be seen in the following diagram. Material aspects if converted into a percentage of 100% valid, construction of 98.8% valid and language by 100% valid.

Diagram 1. Percentage results of validation instrument improvement on Daily Assessment of both validators



After the validation process, cognitive level grouping is performed again. Obtained results dominated in the category of C4 and C5 problems. Where as many as 76% belong to the category of HOTS. Only 24 questions are categorized as LOTS. It can be concluded that the quality of the problem has improved.

Table 8. Grouping cognitive levels of improvement on Daily Assessment

| IPK | ΣQuestion | Cognitive Level | | | | | |
|-------|-----------|-----------------|----|----|----|----|----|
| | | C1 | C2 | C3 | C4 | C5 | C6 |
| 3.2.1 | 10 | - | 1 | 5 | 4 | - | - |

| | | | | | | | |
|--------------|----|---|---|----|----|----|---|
| 3.2.2 | 11 | - | - | - | 7 | 4 | - |
| 3.2.3 | 3 | - | - | 1 | 1 | 1 | - |
| 3.2.4 | 5 | - | - | - | 3 | 2 | - |
| Σ | 25 | - | 1 | 6 | 15 | 7 | - |
| Percentage % | | 0 | 4 | 20 | 51 | 25 | - |

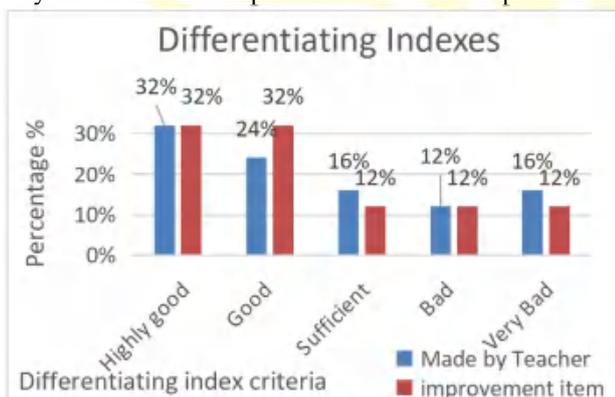
The results of the trial were limited to 30 students processing data from aspects of reliability, differentiating index, difficulty level and distractor. Here are the results of the multiple choice question based on reliability aspects that have been processed in the form of tables.

Tabel 9. Reliability of Teacher-Made item question and the Improvements of Item Question

| | Teacher-Made item question | the Improvements of Item Question |
|--------------|----------------------------|-----------------------------------|
| Reliabilitas | 0,64 | 0,74 |

Based on the reliability table made by <0.70 with not reliable criteria. While the improvement of the item question > 0.70 reliability of can be said to be reliable.

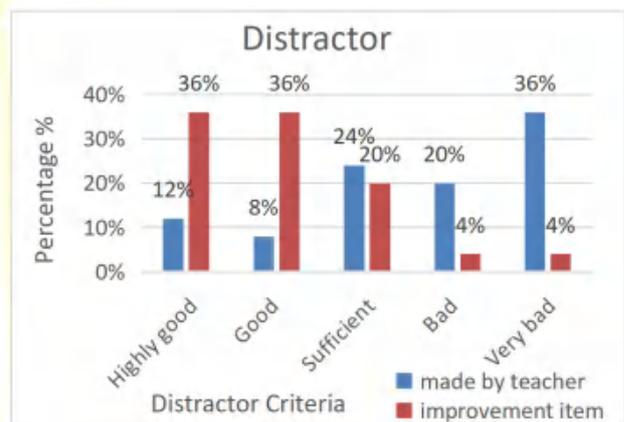
Diagram 2. Results of processing data multiple choice aspects of the differentiating index on the problem made by the teacher and improvement of the item questioned



On the improvement of the question of multiple choice aspects of the differentiating index as many as 8 questions with very good criteria, 8 questions with good criteria, 3 questions of sufficient criteria, 3 questions of bad criteria, and 3 questions of very bad criteria. In the teacher-made questions as many as 8 very good criteria questions, 6 questions of good criteria, 4 questions of sufficient criteria, 3 questions of bad criteria, and 4 questions are very bad. Can be seen in the diagram there is an increase in numbers with good criteria of 8 percent. Accompanied by a decrease in numbers on sufficient criteria and very bad criteria by 4 percent. This means that the improvement of item questions can distinguish

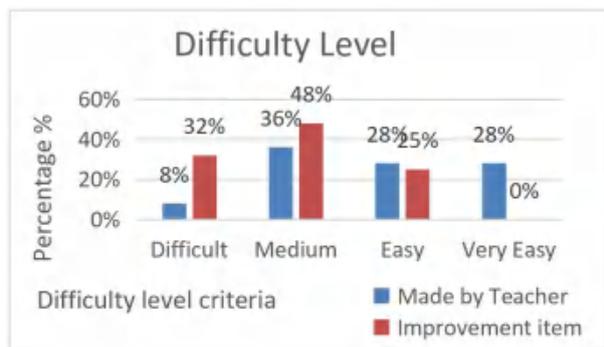
groups of clever students with weak groups of students well. In accordance with the function of differentiating index that is able to distinguish clever groups and weak groups (Arif, 2015). In addition, it can be concluded that the quality of the problem has improved. Aside from the differentiating index aspect. Also done data processing from the aspect of effective distractor.

Diagram 3. Results of data processing multiple choice distractor aspects on the problem made by teachers and improvement of the item questioned



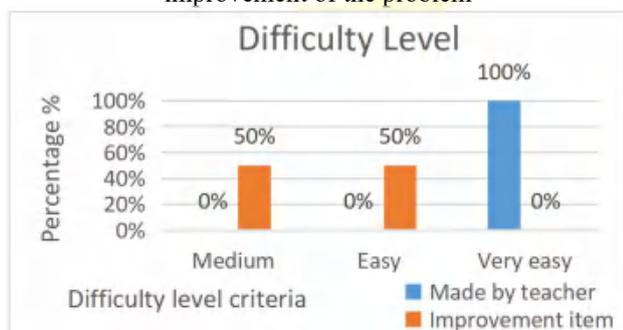
On the improvement of the question of multiple choice distractor aspects as many as 9 problems with highly good criteria, 9 good category questions, 5 enough category questions, 1 problem is bad, and 1 problem is very bad. In the teacher-made questions as many as 3 very good category questions, 2 good criteria questions, 6 questions of sufficient criteria, 5 questions of bad criteria, and 9 questions on very bad criteria. So on the criteria of very good there is a significant increase from 12% to 36%, as well as on good criteria there is a very significant increase from 8% to 36%, accompanied by a decrease in the criteria is sufficient, bad, and very bad. This means that the deceptiveness of the item questions fixes works effectively and there is no misleading deceptive deceptiveness. Misleading deceiving is if the number of clever student groups prefers outwits over the number of weak students who choose the decept (Arif, 2015). Here are the results of data processing aspects of difficulty

Diagram 4. the result processing multiple choice aspects of difficulty level on the problem made by the teacher and improvement item questioned



On the improvement of the question of multiple choice aspects of difficulty level as much as 8 difficult criteria questions, 12 questions of moderate criteria, and 5 questions of easy criteria. In the teacher-made questions as many as 2 difficult criteria questions, 9 medium criteria questions, 7 easy criteria questions, and 7 criteria questions are very easy. Based on the table of easy problem comparison: medium: difficult made by teachers is 56% : 36% : 8%. The problem made by teachers is still dominated by easy categories, not meeting and still far from the proportion of good difficulty levels. The proportion of good difficulty is 1:2:1 or 25% : 50% : 25% (Fitriatun and Sukanti, 2013). After the problem is fixed starting from content, language, construction, materials obtained a ratio of the proportion of difficulty level of 25% : 48% : 32%. Comparison of problem improvements is close to the proportion of good difficulty levels. In addition to multiple choice there are also 4 essays that are done analysis, repair and data processing. Here's a diagram of essay questions based on aspects of difficulty level.

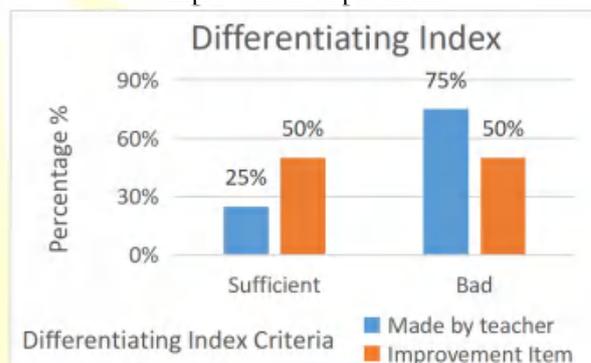
Diagram 5. The result data processing essay aspects of difficulty level on the problem made by the teacher and improvement of the problem



At the difficulty level of essays made by teachers as many as 4 out of 4 questions have very easy criteria with cognitive levels 2 questions C2 and 2 questions C3. So there is no HOTS problem in the essay section made by the teacher. Then made improvements to the question

obtained 2 easy criteria and 2 questions of moderate criteria. With 3 problems are at the cognitive level of C4 and 1 problem at the cognitive level of C5. 4 essay questions after repair fall into the category of HOTS. This is in accordance with the demands of the 21st century about the need for HOTS in training students to think critically, creatively and accustomed to problem solving (Retnoasih, 2018). Here's a diagram of essay. Problems based on the differentiating power aspects

Diagram 5. The result processing essay aspects of differentiating index on problems made by teachers and improvement of problems



There was a significant increase in the criteria by 25% and a decrease in bad criteria by 25%. So that the quality of the problem of experiencing improvement.

Acknowledgment

Researchers express gratitude to the examiner lecturer Dr. Yuliani, M.Si and Dr. Rinie Pratiwi Puspitawati, M.Si; validator Dr. Sifak Indana, M.Pd. In addition, the researchers also expressed their gratitude to Dra. Rubiah S.Pd. as biology teachers and students of grade XI IPA 3 SMAN 20 Surabaya, along with other parties who have been involved and helped so that research activities can be completed.

CLOSING

Conclusion

Daily assessment questions of function material and plant tissue structure made by teachers can be concluded to have low reliability, low difficulty level, good difference index, and bad distractor. Improvement item can be concluded to have high reliability, medium difficulty level, highly good difference index, highly good distractor.

Suggestion

It is recommended to do analysis of question items before being tested directly to students, so that the quality of the

question can be known and can measure the achievement of learning objectives.

REFERENCES

- Arif, M. (2015). Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda. *EduTic - Scientific Journal of Informatics Education*, 1(1), pp. 1–9.
doi: 10.21107/edutic.v1i1.398.
- Arifin, Z. (2011) *Evaluasi Pembelajaran Prinsip, Teknik, Prosedur*. Bandung: PT Remaja Rosdakarya.
- Arifin, Z. (2017) 'Kriteria Instrumen Dalam Suatu Penelitian', *Jurnal Theorems (the original research of mathematics)*, 2(1), pp. 28–36.
- Fitriatun, A. and Sukanti (2013) 'Analisis Validitas, Reliabilitas, Dan Butir Soal Latihan Ujian Nasional Ekonomi Akuntansi Di MAN Maguwoharjo', *Jurnal Kajian Pendidikan Akuntansi Indonesia*, 1(3), pp. 1–11.
- Haryanto (2020) *Evaluasi pembelajaran; Konsep dan Manajemen*. 1st edn, UNY Press. 1st edn. Edited by S. Amalia. Yogyakarta. Available at: <http://staffnew.uny.ac.id/upload/131656343/peneliti-an/Evaluasi Pembelajaran.pdf>.
- Himmah, W. I. (2019) 'Analisis Soal Penilaian Akhir Semester Mata Pelajaran Matematika Berdasarkan Level Berpikir', *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, 3(1), p. 55.
doi: 10.31331/medivesveteran.v3i1.698.
- Iskandar, A. and Rizal, M. (2018) 'Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP', *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(1), p. 12.
doi: 10.21831/pep.v22i1.15609.
- Istika, N., Zamsir, Z. and Prajono, R. (2019) 'Kualitas Tes Ulangan Akhir Semester Mata Pelajaran Matematika Kelas VII Semester II SMP Negeri 12 Kendari', *Jurnal Penelitian Pendidikan Matematika*. doi: 10.36709/jppm.v7i2.8274.
- Kurniawan, D. D. (2019) Analisis Butir Soal Ujian Akhir Semester Matematika Dengan Teori Rspn Butir, *Briliant: Jurnal Riset dan Konseptual*. doi: 10.28926/briliant.v4i2.316.
- Novianti, N. and Siburian, T. A. (2018) 'Telaah Soal dan Analisis Tingkat Kesukaran Butir Soal Ujian Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas VII SMP Swasta Mulia Medan Tahun Pembelajaran 2016/2017', *Kode: Jurnal Bahasa*. doi: 10.24114/kjb.v6i3.10830.
- Prasetyo, N. A. 2013. Analisis Butir Soal Ulangan Akhir Semester Gasal Biologi SMA Kelas XTahun Pelajaran 2012/2013 di Kabupaten Pasuruan dengan Menggunakan program ITEMAN dan kesesuaiannya dengan tuntutan KTSP. Skripsi tidak diterbitkan. Malang: FMIPA Universitas Negeri Malang.
- Rahayu, R. and Djazari, M. (2016) 'Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi', *Jurnal Pendidikan Akuntansi Indonesia*, 14(1). doi: 10.21831/jpai.v14i1.11370.
- Retnoasih, N. (2018) 'Implementasi Pembelajaran HOTS (Higher Order Thinking Skill) IPA Menggunakan Alat Sederhana', *Jurnal Penelitian Pendidikan Matematika Dan Sains (JPPMS)*, 2(2), pp. 48–53.
Available at: file:///C:/Users/HP/Documents/7190-21167-1-SM.pdf.
- Riduwan (2013) *Skala Pengukuran Variabel-Variabel Penelitian*. Bandung: Bandung : Alfabeta.
- Sudjiono, A. (2012) *Pengantar Evaluasi Pendidikan*. 15th edn. Jakarta: PT Raja Grafindo Persada.
- Sukmela, Y. P. and Nofrion (2018) 'Analisis Butir Soal Paket A Ujian Akhir Semester Genap. *Jurnal Buana*, 2(5), pp. 467–479. Available at: <http://geografi.pjj.unp.ac.id/index.php/student/article/view/106>.
- Supiyansyah, H., Kusumah, I. H. and Berman, E. T. (2017) 'Analisis Kualitas Soal Ulangan Akhir Semester Genap pada Mata Pelajaran Produktif Program Keahlian Teknik Kendaraan Ringan', *Journal of Mechanical Engineering Education*. doi: 10.17509/jmee.v4i1.7441.
- Wijayanto, P. A., Allifah, A. and Amirrudin, A. (2016) 'Evaluasi Kualitas Instrumen Tes Dalam Pembelajaran Geografi di MAN 2 Kota Batu', *Jurnal Geografi*.
doi: 10.15294/jg.v13i2.7969.
- Yuniar, M., Rakhmat, C. and Saepulrohman, A. (2015) 'Analisis Hots (High Order Thinking Skills) Pada Soal Objektif Tes Dalam Mata Pelajaran Ilmu Pengetahuan Sosial (Ips) Kelas V Sd Negeri 7 Ciamis', *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 2(2), pp. 187–195. Available at: <https://ejournal.upi.edu/index.php/pedadidaktika/article/viewFile/5845/3961>.
- Zamzania, W. H. and Aristia, R. (2018) 'Jenis-Jenis Instrumen dalam Evaluasi Pembelajaran', *Universitas Muhammadiyah Sidoarjo*.