

# The Effectiveness of the Zipgrade-Assisted Learning Outcomes Assessment Analysis in Promoting Indonesian Vocational Teachers' Competence

Suhendar<sup>a</sup>, Herman Dwi Surjono<sup>b</sup>, Slamet PH<sup>c</sup>, Priyanto<sup>d</sup>, <sup>a,b,c,d</sup>Yogyakarta State University, Yogyakarta, Indonesia, Email: <sup>a</sup>[suhendar@untirta.ac.id](mailto:suhendar@untirta.ac.id), <sup>b</sup>[hermansurjono@uny.ac.id](mailto:hermansurjono@uny.ac.id), <sup>c</sup>[slamet.ph@uny.ac.id](mailto:slamet.ph@uny.ac.id), <sup>d</sup>[priyanto@uny.ac.id](mailto:priyanto@uny.ac.id)

One of the tasks of Indonesian vocational school teachers is to correct the assessment of learning outcomes and analyse them. The assessment activities that must be carried out by teachers are performance assessments, self-assessments, portfolio assessments, daily assessments, midterm assessments, end of semester assessments, competency level exams, and school exams. For the majority of teachers, these activities are too administrative because the results must be corrected and analysed. The teacher makes the questions and corrects the assessment of learning outcomes still using the manual method. In fact, it is very rare for teachers to analyse the results of their assessment. Therefore, the research objectives are: 1) to improve teacher competence in analysing the learning outcomes of Indonesian vocational students; 2) to make effective the process of the ZipGrade-assisted learning outcomes assessment. Quasi-experimental, the one-group pretest-posttest design was adopted to complete the research. The survey questionnaire was designed for exploring the perceptions of teachers and students the effectiveness of using ZipGrade in the learning outcomes assessment. The findings revealed that the teacher had positive experiences about how to analyse the assessment of learning outcomes. The teacher is greatly helped by ZipGrade to effectively correct student answers. Students argue that the using of ZipGrade-based android answer sheets has simplified the filling process. The teachers explained that the analysis of ZipGrade student learning outcomes assessment made their work easier. ZipGrade has helped him/her by making correcting students' answers very easy and effective. In addition, ZipGrade provides information on the results of the validity analysis, the level of difficulty of the questions, and the

distinguishing features of each item. In the future, teachers hope to get training in the use of other technologies to compile and evaluate learning outcomes.

**Key words:** *Effectiveness, ZipGrade, Learning Outcomes, Assessment, Teacher's Competency, Vocational.*

## Introduction

The teacher and lecturer law states that the teacher is a professional educator with the main task of teaching, guiding, directing, training, assessing, and evaluating of students (Undang-Undang No. 14 Tahun 2005 tentang Guru dan Dosen, 2005). One of the tasks of the teacher is to assess (Sudarwanto & Sulistyowati, 2019). Educational assessment is the process of gathering and processing information to measure student learning outcomes (Ismail et al., 2019). Assessment of student learning outcomes includes performance assessments, self-assessments, portfolio assessments, daily assessments, midterm assessments, end of semester assessments, competency level exams, school exams, and national exams. The Directorate of PSMK has compiled an assessment guide as a practical reference for teachers in planning and assessing the learning outcomes of vocational students (Tim Dikdasmen, 2018). This guide is a practical guide for processing and making reports of the results of the assessment in an accountable and informative manner.

At SMK Indonesia, assessment has been carried out throughout the learning process. Assessment is used as a basis for improving the learning process. Writing is used as a learning process that allows students to be involved in the assessment process (Depdiknas, 2008). The Assessment provides opportunities for students to improve their learning outcomes so that there is more leverage (Baird et al., 2014). At the end of the lesson, an assessment is given to measure the achievement of competencies. The scope of assessment in SMK, includes principles, mechanisms, procedures, techniques, instruments, processing, and follow-up assessment and reporting of student competency achievements. The purpose of assessing student learning outcomes at SMK in Indonesia is to determine the level of student learning outcomes, to determine student growth and development, to diagnose student learning difficulties, to find out the effectiveness of the learning process, and to determine curriculum achievement (Prieto et al., 2019). Assessment is one of the teacher's primary duties in the learning process. Assessment in education is one aspect of the competency assessment process (Mukhtar & Ahmad, 2015).

Furthermore, the assessment of learning outcomes at primary and secondary education levels is carried out by educators, education units and the government. There are various ways to assess the results of the student learning experience; the most common methods include using

demonstrations, written tests, observations, and interviews (Santosa & Dwi, 2018) including the use of technological devices and computerised scoring (Hursen, 2011) (Kush et al., 2012). The development of digital technology today has supported many teaching, learning and assessment activities. Although on the other hand, these rapid technological changes have caused social inequalities in their use (Wikström & Eklöf, 2019). The development of integrated information and communication technology (ICT) competencies with education is very effective in increasing teacher competency (Demchenko, 2018).

The teachers are not only competent in social media but the teachers must understand and utilise technological resources for learning activities in the classroom. Even until now, most teachers don't analyse the questions used for assessment. Therefore, they don't know the validity and reliability of the questions they make. According to the facts obtained, the teachers find it difficult to carry out such an analysis. This is why they need a technology that can make it easier for them to do the analysis. Teachers' skills could improve in analysing cognitive test assessment results using the ZipGrade base on smartphones (Castillo García et al., 2019). Teacher skills are developed to include the skills to compile effective test questions, the skill of correcting test questions using a smartphone, and the ability to analyse items.

It is not enough for the teacher just to ask questions, test them, and correct students' answers. The teacher must also be able to correct the questions they made and analyse the results of their assessment. During an assessment, the teacher needs to arrange and develop test questions, as well as analyse the results. An analysis is performed through gathering, summarising, and using the information obtained from student tests results in the assessment (Cherner, T., Lee, C-Y., Fegely, A., & Santaniello, 2016). This means that the analysis is carried out to study and examine each test item to obtain quality questions that can be used several times. Also, the analysis is to improve the quality of test items, revisions or elimination of ineffective questions. Besides, analysis can also be used to ascertain the level of student understanding of the material. It is important to use quality questions because they can provide accurate information about how effective the learning strategy is, and also determine the level of student understanding (Geisinger, 2019).

Analysis of items can either be conducted quantitatively or qualitatively; while the quantitative determines the characteristics of the items, the qualitative determines the relationship of content and shape (Bichi, 2015). This is useful for improving judgment and procedures empirically. Also, the qualitative analysis helps in the construction and validation of content, while quantitative analysis measures the difficulty of the items by checking the validity and reliability of the questions (Quaigrain & Arhin, 2017). Therefore, in special assessments, the use of validity is very necessary.

The questions made by each teacher must be validated by experts and peers, because validation of questions is also very important before being tested on students. Validity is an evaluative policy that is integrated with empirical facts and theoretical reasons that support the adequacy and appropriateness of inference and action based on test scores. It helps in the interpretation of test scores and provides measurement accuracy. Reliability is the degree of consistency of two measurement scores of the same object using different measuring instruments and scales (Reynolds, C., Livingstone, R.B., Willson, 2010). This is why student learning outcomes can be called reliable if results from different tests are the same or different items share the same characteristics. The test is said to be reliable if the observation score had a high correlation with the actual score (Geisinger, 2019). Therefore, reliability is the correlation coefficient between two observation scores gotten from the measurement results using a parallel test. This means that a test can be called reliable if the measurement results meet the set of conditions already mapped out by the test taker.

The teacher will get knowledge about how to analyse the difficulty level of the questions and the different power of the items. According to (Quaigrain & Arhin, 2017), the item difficulty index should be at intervals of 0.3– 0.7. At this interval, information about students' abilities will be obtained maximally. Furthermore, the index differentiation on any item is said to be good if it is greater or equal to 0.3. This is because the small differentiation index of an item is not able to distinguish the ability of students. Also, when using Content-Referenced Measures in test analysis, the item index differentiation is not considered, as long as it is not negative (Sudha & Janardan, 2014). The low value of the index differentiation decreases the validity of the test. The correct answers were given a score of 1, while incorrect answers were given a score of 0. The item difficulty index (P) is divided into 3 categories i.e. easy ( $P > 0.70$ ), Moderately ( $0.31 \leq P \leq 0.70$ ), and Difficult ( $P \leq 0.30$ ) (Bichi, 2015). The following equation was used to find difficulty index (P) (Boopathiraj & Chellamani, 2013).

$$P = \frac{Ru + RI}{Nu + NI}$$

Where,

Ru = the number students in the upper group who responded correctly

RI = the number students in the lower group who responded correctly

Nu = Number of students in the upper group

NI = Number of students in the lower group

The discrimination index scores for each item were sorted from the lowest to the highest. After that, the distribution of scores was then divided into two groups (lower and upper groups). The lower group consisted of participants with low scores and the Upper group consisted of participants with high scores. Furthermore, the data in the table was then used to determine the number of participants in the upper and lower groups (A and B) and the number of participants in the upper and lower groups (AB and BB) who answered correctly.

The discrimination index (D) was calculated using the following equation (Boopathiraj & Chellamani, 2013).

$$D = \frac{R_u - R_l}{N_u \text{ (or) } N_l}$$

Test item analysis were carried out to identify test deficiencies. Furthermore, item analysis also helped to assist evaluation, compile informal and local tests, create effective items, improve tests materially and increase the validity of questions and reliability (Yıldız & Fidan, 2019). Item analysis is also useful for determining the appropriateness of item functions, providing input to students' ability discussions, providing information about student difficulties, developing a curriculum based on certain aspects, revising the material being assessed, and improving question writing skills (Resnick & Schantz, 2017).

## Research Method

The quasi-experimental with the one-group pretest-posttest design was adopted to finish the research. The participants of this study were the 30 teachers and their students of productive subjects at SMKN 1 Jakarta. A short training was given to improve effective teacher skills (Kandiller & Özler, 2015). The instruments used to collect data were a pretest and a posttest applied to the experimental group (Cuesta et al., 2019). Teachers can easily find out the validity, reliability, convenience, and differentiation of test items through the use of ZipGrade. Research design is presented in table 1 below.

**Table 1:** The quasi-experimental with the one-group pretest-posttest research design

Group	Pre-test	Treatment	Post-test
Experiment	$O_1$	$X$	$O_2$
<b>(P)</b>			

$O_1$  = pre-test carried out before treatment.

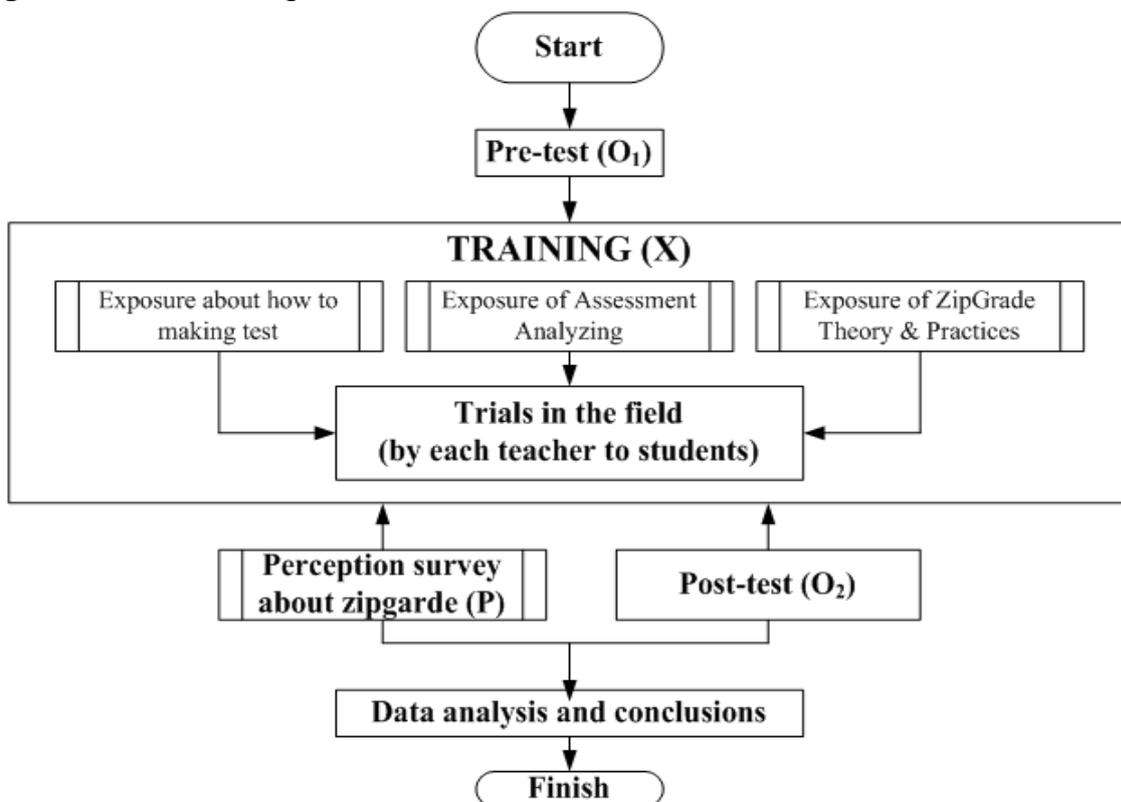
$O_2$  = *post-test* carried out after treatment

$X$  = *teachers were given treatment in the form of using ZipGrade*

P = Perception survey about ZipGrade (Y)

A survey questionnaire was designed for exploring the perceptions of teachers and students about the effectiveness of the ZipGrade-assisted learning outcomes assessment analysing. Data and findings have been obtained by designing research as described in the following figure.

**Figure 1.** Research design



The teacher as a participant is given a pre-test ( $O_1$ ) to answer a number of questions about how to make questions, analyse assessments, and use the student's answer selection tool. In the next step, the teachers are given a short training. Training as a treatment ( $X$ ) was given to improve teacher competency in making questions, validate, test their reliability, and use ZipGrade to correct and analyse the assessment of learning outcomes. After the teacher gets the training, then the teacher is given time to try out the use of ZipGrade. The test was conducted to correct student test answers in their respective subjects. Teacher competency improvement was measured through a post-test ( $O_2$ ). In addition, this study also explored the perceptions of teachers and students about the use of ZipGrade. Pre-test data, post-test, and survey ( $Y$ ) results were analysed to draw a conclusion.

## Research Findings and Discussion

### *The Result of the Pre-Test*

The study began with a pre-test to find out how far the achievement of teacher competencies about the assessment process went. The questions that were asked in the pre-test included how to make effective test questions, analyse the results of the assessment, and use technology as a tool to include ZipGrade. Pre-test results will be a reference for developing

training in the assessment of learning outcomes and the use of ZipGrade as a tool for assistance. Pre-test results are briefly explained in the following figure.

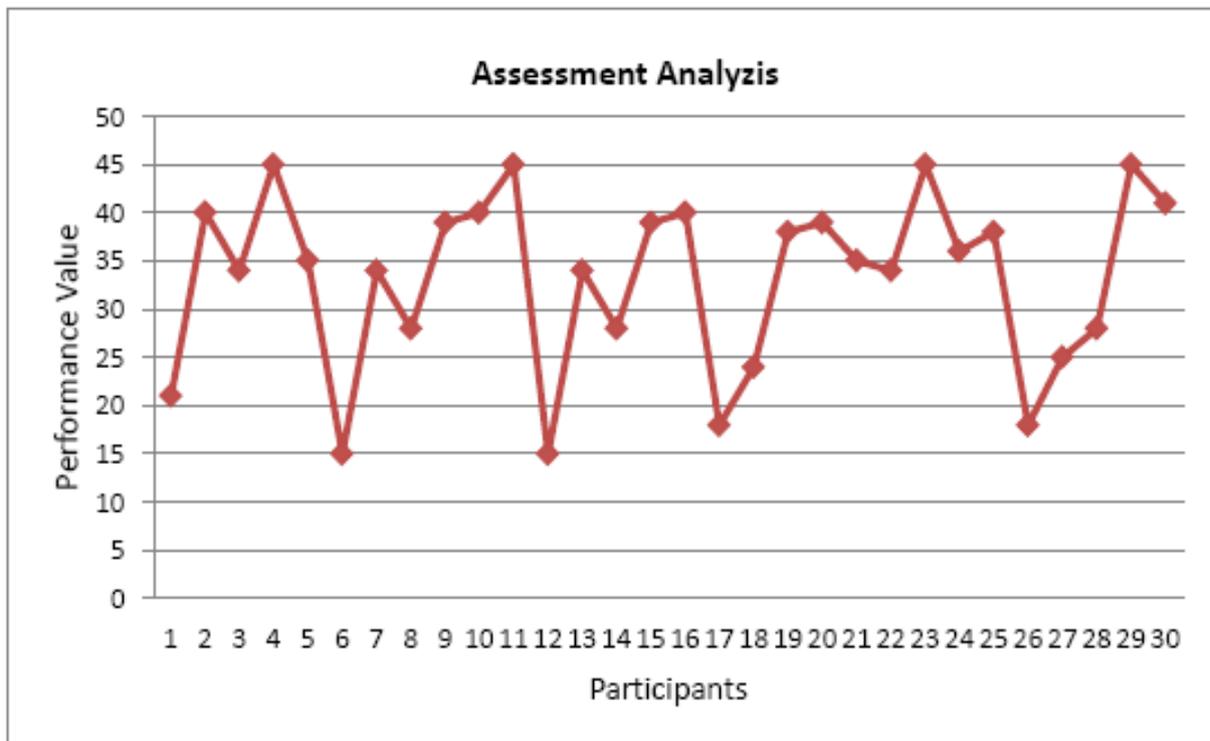
In figure 2, the pre-test results indicate that the achievement of participants on how to make effective test questions is still low. The highest score is 45 and the lowest value is 11. Teachers' understanding of making quality test questions is still low and unequal. This gives an indication that in competency, the teacher must still be given training.

**Figure 2.** The chart of pre-test result of the questions of how to make test item



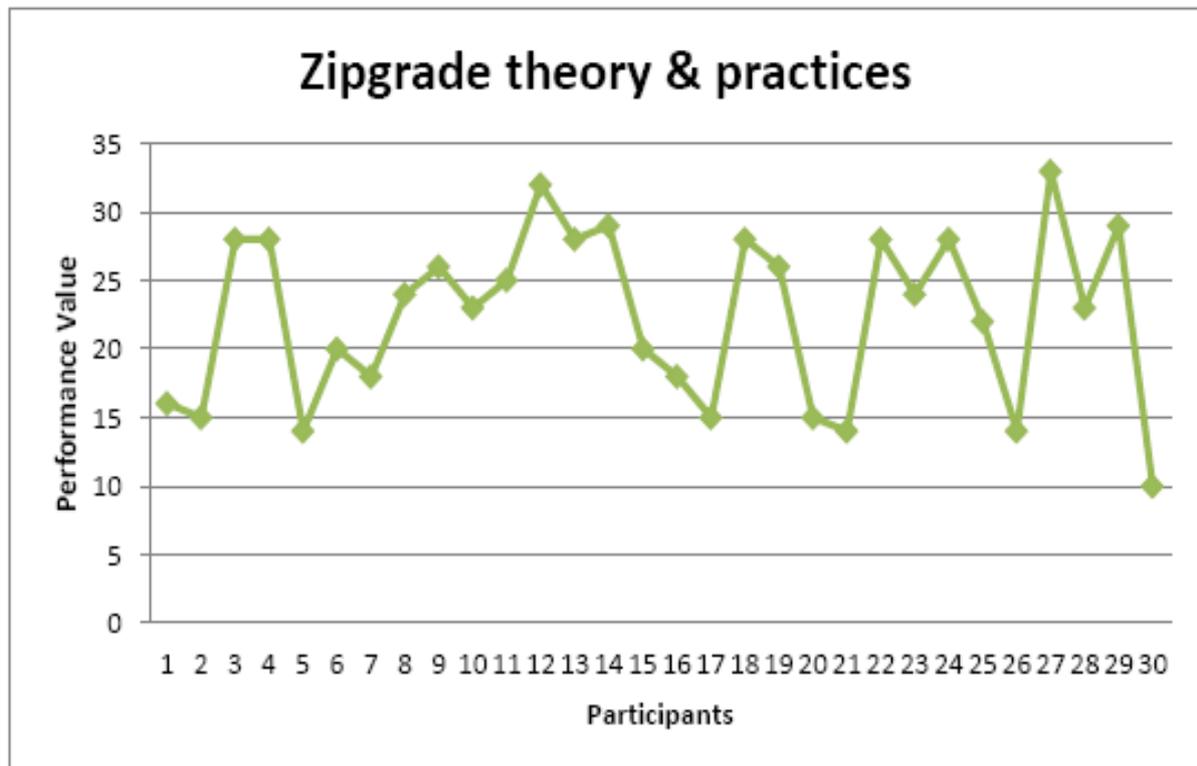
Thirty teachers were tested to answer questions about how to do an assessment analysis. Questions are submitted via practice. The questions were in the form of multiple choice. Pre-test questions totalled 5 items. The results of the practice show that the ability of teachers to analyse the assessment of learning outcomes is still very low. The highest score obtained by the teacher in the pre-test is 45 and the lowest score is 15, as shown in the graph in figure 3. So far the teacher is accustomed to making questions and testing them to students but rarely does an analysis of the questions made and obtain the final results. (Sudha & Janardan, 2014).

**Figure 3.** The chart of pre-test result of the questions of the assessment analysis



Even teachers rarely use technology tools as an aid for analysing learning outcomes. Some schools provide a computer-aided analysis system but not teachers who use it directly. Teachers only accept test scores in the form of final scores without getting an explanation of the overall analysis. Likewise, knowledge about ZipGrade and its use is still not known in depth by the teacher. Some questions about ZipGrade become a part that is tested on the teacher in the pre-test. The results of the practice obtained are far lower than the 2 groups of practice questions before. Achievement of practice results is shown in the graph in figure 4. The lowest value is 10 and the highest value is 32.

**Figure 4.** The chart of pre-test result of the questions of ZipGrade theory & practices



Differences in the achievement of vocational school teacher scores in the three categories of questions: how to make a test item, assessment analysis, and ZipGrade theory & practices as shown in the following table 2.

**Table 2:** Recapitulation of pre-test scores

No	The scope of the pre-test questions	Maximum Value	Minimum Value
1	How to make test item	45	11
2	Assessment analysis	45	15
3	ZipGrade theory & practices	32	10

Referring to the pre-test results as described, illustrated, and summarised in the table above provides recommendations that teachers need to be given training. The training is prioritised to improve teacher skills in analysing ZipGrade assisted learning outcomes.

### ***The Training Process of ZipGrade***

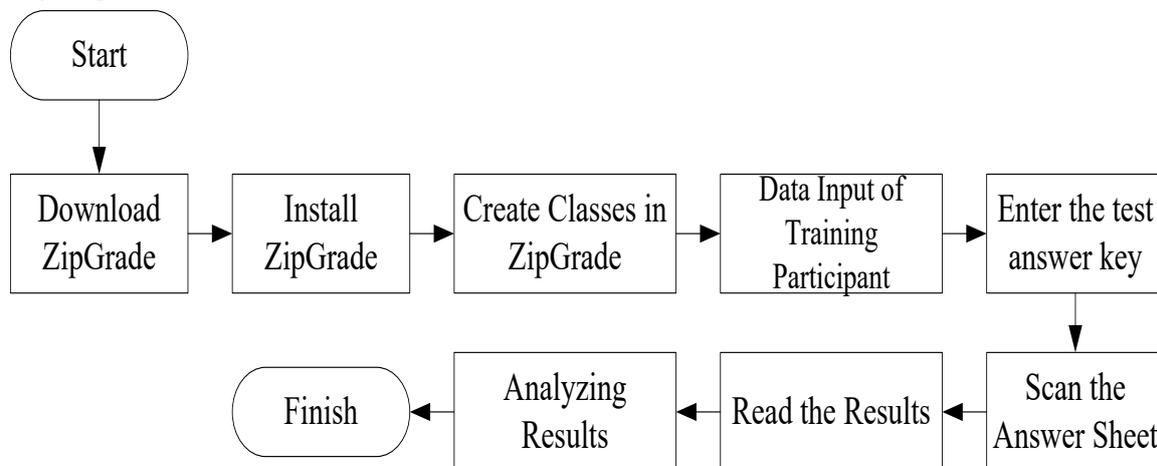
The subjects in this study were vocational school teachers. The teachers were trained to analyse the results of the multiple-choice questions used in cognitive test assessment. They

did this using the ZipGrade application. This application can correct answer sheets and produce relatively better analysis results. Also, ZipGrade can detect answers written using pens, pencils, and markers (Anoname, 2019). The application can increase the competence of teachers in analysing the results of assessments.

ZipGrade is an application for correcting android answer sheets (LJA). LJA makes it easy for students in answers by blackening the circle of choices using a pen, pencil, or marker. The difference between the two lies in technology; LJK needs a scanner and PC/laptop while LJH only needs an Android Smartphone or iPhone. The stages of developing teacher skills in analysing cognitive test scores are shown in Figure 1. The teachers were given a pre-test on the theory of the process of compiling questions, statistical analysis, test item difficulty, differentiation analysis, and the technology used in the process. The test instrument in this study relates to the material presented in the short training.

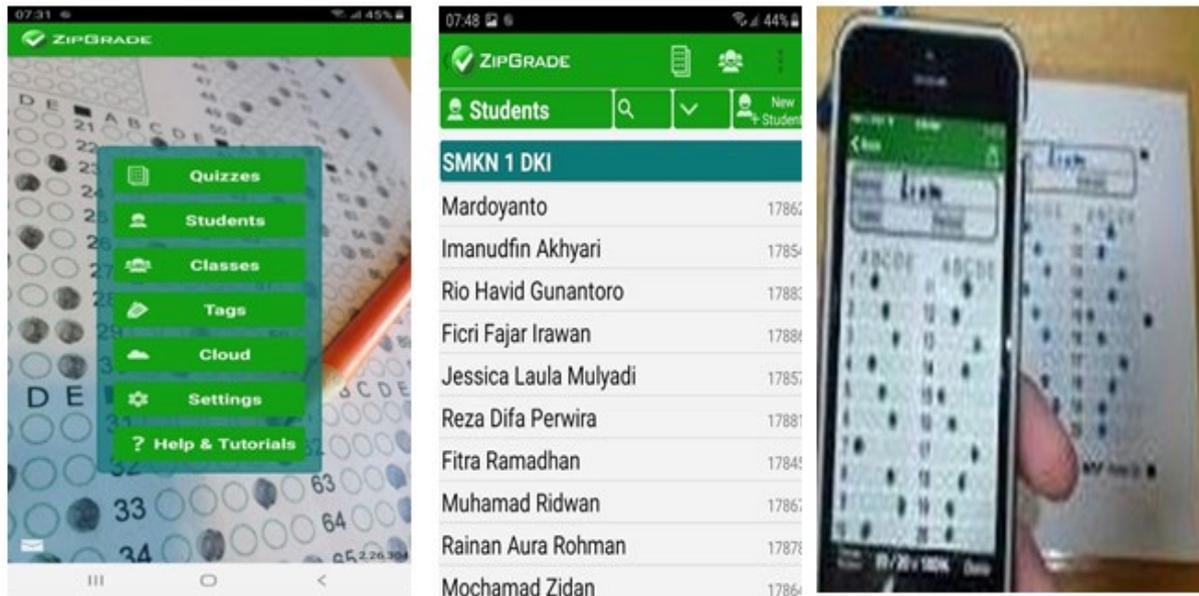
In the practice phase (see Figure. 1), each participant was guided to use ZipGrade starting from downloading, installing, and creating classes. See the following Figure 5.

**Figure 5.** Stages of the practice of using ZipGrade to improve teacher competence in analysing the results of assessments



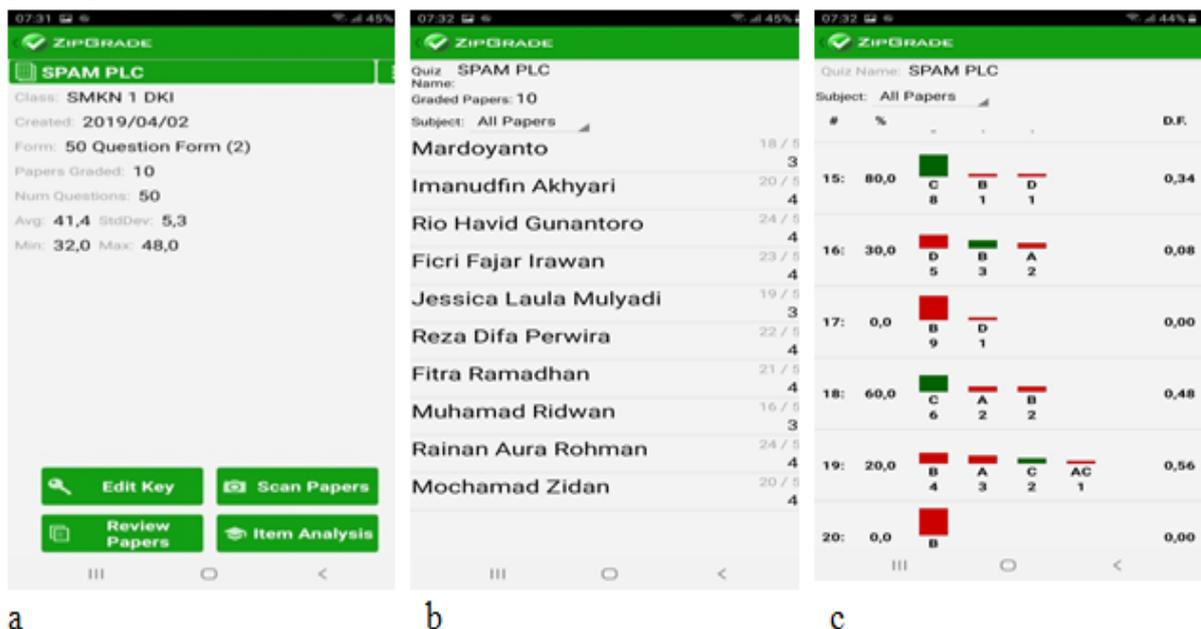
Participants (the teachers) were also guided to input name and answer key, correct answers, and analyse the results into the smartphone as in the following figures.

**Figure 6.** Three stages of preparation set ZipGrade for the correction process



In Figure 6a, the teacher does 3 main things, namely determine the type of quiz, enter student data (test participants), and make the class easy. The three main steps that have been taken will form the results shown in Figure 6b. In this picture, the teacher can see the class name and list of students in the class. Other classes and names of students just need to be added as the process in Figure 6b. The teacher easily corrects the answer fields in the circle of available options by scanning it using an android smartphone (Figure 6c).

**Figure 7.** ZipGrade data presentation for analysis of student assessment results (test participants)



The step of scanning students' answers will produce Figure 7 that is ready to be analysed by the teacher. The teacher can see and interpret the results of each student's answer. In addition to being able to see the value of each student, the teacher can analyse the overall results through the display shown in Figure 7c. The teacher can see the level of difficulty for each item. The number of students who answered correctly for each item and its distinguishing factors can be analysed by the teacher by viewing the ZipGrade (see Figure 7c). Furthermore, a teacher competency improvement training on item-level analysis was carried out using ZipGrade and a Smartphone. After that, the difficulty index was calculated based on the corrections that were made to the results of student answer sheets.

### ***The Results of the Post-Test***

At the end of the research process, the teacher was given a post-test. The type and substance of the contents of the post-test questions were the same as the pre-test questions. A short training course on how to write quality questions, how to analyse learning outcomes assessment, and the use of ZipGrade as a tool has helped improve teaching skills. Teachers who have been given training have improved knowledge and skills. This is evidenced by the results of the post-test he/she has achieved. Drastic improvement in the value of the post-test results compared to the pre-test shows that in general, the teacher can easily receive additional knowledge and insights related to the improvement of his components.

As can be seen from the average value of the normalized gain, the teacher's understanding of the assessment results analysis increased after the training. Table 3 shows the P-value of pre-test, post-test, and normalized gain values.

**Table 3:** P-value recapitulation of teacher understanding improvement

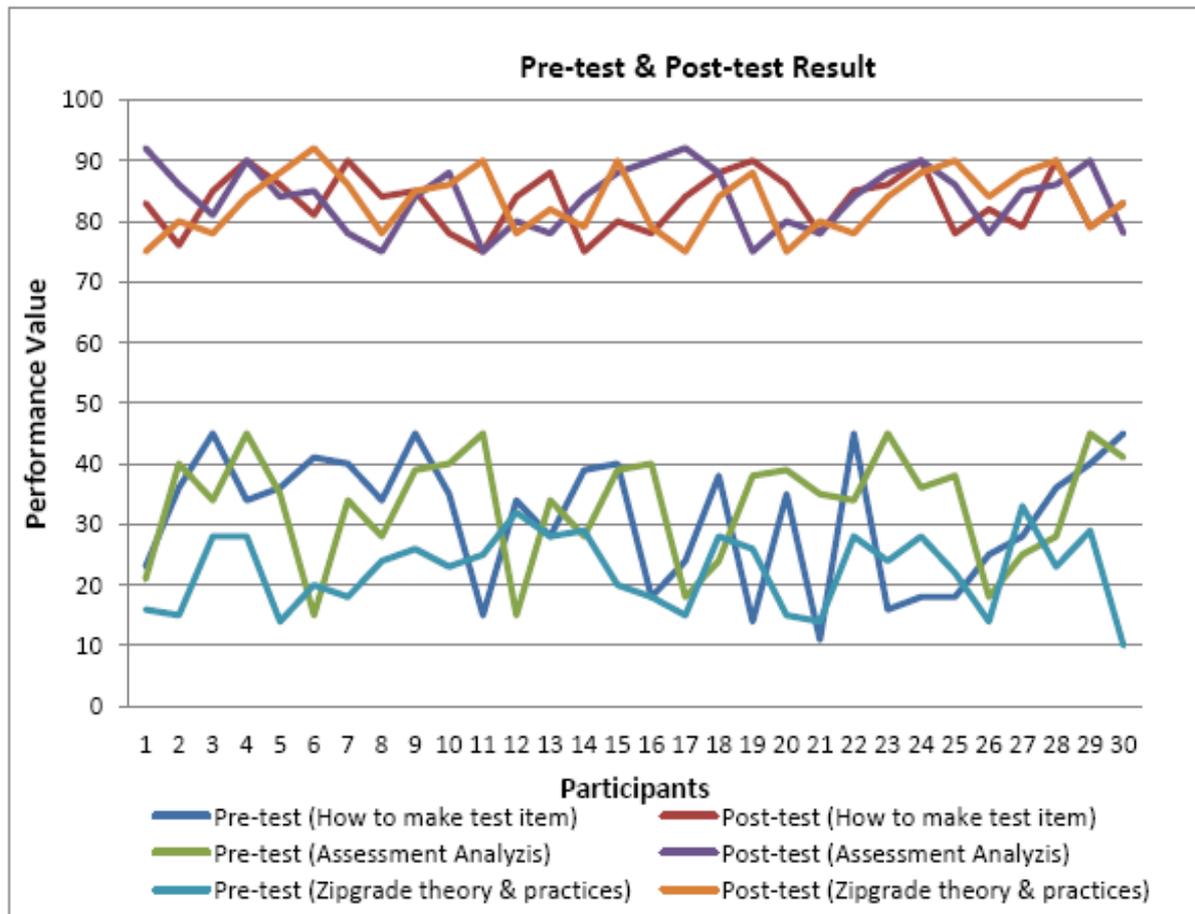
Test	Ideal scores	Max Score	Min score	<G>	<g>	
					Value	Criteria
Pre-test	100	45	10	42	0.74	High
Post-test	100	92	75			

The normalised gain value is 0.74. This figure shows an increase in the teachers' understanding of the training theme. The criteria for improvement in the high category indicates that training can improve teachers' understanding of the use of ZipGrade in analysing assessment results. This means that if teachers get accustomed to using ZipGrade, that category can change to high. Increased intelligence will improve the quality of work and increase also the desire to learn and achieve (Lottridge et al., 2018).

The training provided was continued with a direct trial by the teacher which proved to be very effective in increasing his/her competence. Teachers better understand and master how

to make questions, analyse assessments, and choose and use technology tools as tools for the examination correction process after they are given time to test them directly on students. The difference in competency achievement before the teacher is given training and after the teacher gets training and given the opportunity to practise it, is shown through the graph in Figure 8 below.

**Figure 8.** The chart of Pre-test & Post-test Result



The training of ZipGrade can improve the ability of teachers. The teachers can analyse the items. The teacher understands how to distinguish the difficulty level of the questions. The teacher can interpret the different power of questions. ZipGrade helps teachers to correct student answers. Thus the assessment process was more effective and efficient (Geisinger, 2019), even though so far the teacher only makes questions and evaluates students' answers. The teacher did not understand how to make quality questions. The teacher did not know how to correct quickly, effectively and efficiently (Hanifah, 2014).



### ***The Results of Perception Survey about ZipGrade***

In Indonesia, not all vocational schools have complete information technology facilities. Every student does not have a cell phone and not all cell phones have good facilities and capacity. Therefore, teachers must find creative ways for learning to proceed well, including carrying out assessment activities effectively. To overcome the incompleteness of this facility, ZipGrade is very effective in being used as a correction tool and analysis of student learning outcomes assessment. Not only is ZipGrade effective for analysing multiple choice objective questions, but ZipGrade can be developed to correct and analyse other objective questions, namely true-false, match-up, and more than one correct answer.

In this study, the perception of students and teachers were explored to obtain information on the extent and effectiveness of ZipGrade as a tool for assessing learning outcomes. Five things were surveyed to get students' perceptions about the effectiveness and practicality of using ZipGrade. In this study, students and teachers' perceptions were explored to obtain information on the extent and effectiveness of ZipGrade as a tool for assessing learning outcomes.

Data has been generated from opinion surveys of students and teachers about the effectiveness of using ZipGrade in the assessment process of learning outcomes. The survey results are explained in the following table.

**Table 4:** The result of a perception survey about the effectiveness of using ZipGrade

Resources of Information	Perceptions
Students	<ol style="list-style-type: none"><li>1) Students fill in ZipGrade answer sheets easily using pencils, pens, or markers</li><li>2) Students can see the test results directly just before leaving the examination room</li><li>3) Answer sheets can be photographed by themselves to anticipate errors in filling in data and answers</li></ol>
Teachers	<ol style="list-style-type: none"><li>1) ZipGrade makes it easy for teachers to quickly correct student test answers</li><li>2) The process of student data input can be done by GRU or assisted by students</li><li>3) The facilities in the ZipGrade application make it easy for teachers to return assessment results to students online through Japri or social media groups</li><li>4) High accuracy for correcting the results of tests anywhere and anytime using a smartphone</li><li>5) ZipGrade application helps teachers to correct and analyse the items they make</li><li>6) The teacher can create a variety of objective objectives: true-false, multiple-choice answers, multiple choices, and match</li><li>7) The teacher can create higher quality questions referring to the ZipGrade assisted assessment results</li><li>8) Student data and assessment results can be stored in a cloud database</li></ol>

In addition to being pedagogically skilled, the teachers must master technology to help their professionalism in teaching and learning (Han et al., 2019). ZipGrade is one of the technological media that teachers must master. Teachers can easily correct ZipGrade assisted. Thus the teaching and learning process at SMK Indonesia will run effectively.

## Conclusion

In this research, the effectiveness of the ZipGrade-assisted learning outcomes assessment analysis to promote the Indonesian vocational teachers' competence was measured in high-stakes tests. From the gathered data it is clear that the teachers who participated reached a better overall result. The teachers and their students who participated in the training are more diligent persons. The training of ZipGrade made it easier to recall the information in the post-test because they had seen it before.



The ZipGrade is more useful for teachers in correcting of multiple-choice questions. ZipGrade aids the teachers in analysing the results of assessments. The application also helps the teachers to create group or grade assessments. Furthermore, students can also help the teachers to input their data into the application. The answer sheets can be scanned using mobile phones and the grade statistics will be shown in less than 5 seconds.

Students find it easy to write answers using android answer sheets. The convenience is caused by students being able to use pencils, pens, or markers to circle the answer choices on the LJA. Furthermore, teachers feel motivated to develop quality multiple-choice and true-false questions. The development of many multiple-choice questions will make it easier for the teacher to correct them as soon as possible. Short training on using ZipGrade is very effective as a tool to correct student assessment results. The teachers no longer need to bring the student answer sheet home to further prove it. When students finish answering all the questions, the teacher can correct them when the students leave the class or test site.

ZipGrade can be used to determine the validity, difficulty level, and differentiation of each test item. ZipGrade can also be used as a tool to answer and correct true-false questions. This application is easier to use than SPSS or ITEMAN because analysis data can be exported to CSV and pdf formats. The application can also share the test results.

### **Acknowledgements**

This publication is dedicated to the Indonesian educational fund management agency (LPDP) and Directorate of higher education, Indonesian Ministry of education and culture who are thanked for their assistance both in information, funding and policy regulation on improving lecturer publications.

## REFERENCES

- Anoname. (2019). ZipGrade. Retrieved January 10, 2019, from <https://zipgrade.zendesk.com/hc/en-us/articles/201168729-How-do-I-get-started-with-ZipGrade>
- Baird, J.-A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the Field Review Assessment and Learning. Norwegian Knowledge Centre for Education*. (13).
- Bichi, A. A. (2015). Item Analysis using a Derived Science Achievement Test Data. *International Journal of Science and Research*, 4(5), 1655–1662.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189–193.
- Castillo García, E., Ibañez Torres, L., Leon Bello, G., Sanahuja Santafé, A., Villagrasa Sebastián, V., & Martínez-Solís, I. (2019). *Zipgrade: herramienta para corrección de cuestionarios de evaluación y análisis de resultados*. 67–75. <https://doi.org/10.4995/inred2019.2019.10403>
- Cherner, T., Lee, C-Y., Fegely, A., & Santaniello, L. (2016). A detailed rubric for assessing the quality of teacher resource apps. *Journal of Information Technology Education*, 15(Innovations in Practice), 117–143.
- Cuesta, P. V. N., Mayorga, J. I. M., Padilla, Y. N. P., & Mayorga, A. P. O. (2019). Alternative Assessment Tools for the Writing Skill Development of EFL Learners. *European Scientific Journal ESJ*, 15(7), 177–187. <https://doi.org/10.19044/esj.2019.v15n7p177>
- Demchenko, I. (2018). Forming of Future Teachers' ICT-Competence: Canadian Experience. *Comparative Professional Pedagogy*, 6(1), 54–60. <https://doi.org/10.1515/rpp-2016-0008>
- Depdiknas. (2008). *Panduan Analisa butir soal*. Jakarta, Indonesia.
- Geisinger, K. F. (2019). Empirical Considerations on Intelligence Testing and Models of Intelligence: Updates for Educational Measurement Professionals. *Applied Measurement in Education*, 32(3), 193–197. <https://doi.org/10.1080/08957347.2019.1619564>
- Han, I., Han, S., & Shin, W. S. (2019). Teachers' and students' perspectives on good teaching using technology in elementary classrooms. *International Journal of Information and Communication Technology Education*, 15(3), 103–116. <https://doi.org/10.4018/IJICTE.2019070108>

- Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda. *Sosio E-Kons*, 6 No.1(Asosiasi Mata Pelajaran Ekonomi), 41–55.
- Hawley, L. R., Bovaird, J. A., & Wu, C. R. (2017). Stability of Teacher Value-Added Rankings Across Measurement Model and Scaling Conditions. *Applied Measurement in Education*, 30(3), 196–212. <https://doi.org/10.1080/08957347.2017.1316273>
- Hursen, C. (2011). Technological proficiency perception assessment of teachers in vocational high schools. *Procedia - Social and Behavioral Sciences*, 28, 977–981. <https://doi.org/10.1016/j.sbspro.2011.11.180>
- Ismail, M. A. A., Ahmad, A., Mohammad, J. A. M., Fakri, N. M. R. M., Nor, M. Z. M., & Pa, M. N. M. (2019). Using Kahoot! as a formative assessment tool in medical education: A phenomenological study. *BMC Medical Education*, 19(1), 1–8. <https://doi.org/10.1186/s12909-019-1658-z>
- Kandiller, S. T., & Özler, D. (2015). From Teacher to Trainer: What Changes? Or Does It? *Procedia - Social and Behavioral Sciences*, 199, 436–452. <https://doi.org/10.1016/j.sbspro.2015.07.530>
- Kush, J. C., Spring, M. B., & Barkand, J. (2012). Advances in the assessment of cognitive skills using computer-based measurement. *Behavior Research Methods*, 44(1), 125–134. <https://doi.org/10.3758/s13428-011-0136-2>
- Lottridge, S., Wood, S., & Shaw, D. (2018). The effectiveness of machine score-ability ratings in predicting automated scoring performance. *Applied Measurement in Education*, 31(3), 215–232. <https://doi.org/10.1080/08957347.2018.1464452>
- Mukhtar, M. I., & Ahmad, J. (2015). Assessment for Learning: Practice in TVET. *Procedia - Social and Behavioral Sciences*, 204(November 2014), 119–126. <https://doi.org/10.1016/j.sbspro.2015.08.124>
- Nitko, A. J. (1996). *Educational Assessment of Students* (Second Edi). Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.
- Prieto, M. C., Palma, L. O., Tobías, P. J. B., & León, F. J. M. (2019). Student assessment of the use of kahoot in the learning process of science and mathematics. *Education Sciences*, 9(1). <https://doi.org/10.3390/educsci9010055>
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1–11. <https://doi.org/10.1080/2331186X.2017.1301013>
- Resnick, L. B., & Schantz, F. (2017). Testing, teaching, learning: who is in charge? *Assessment in Education: Principles, Policy and Practice*, 24(3), 424–432. <https://doi.org/10.1080/0969594X.2017.1336988>



- Reynolds, C., Livingstone, R.B., Willson, V. (2010). *Measurement and education*. New Jersey: Pearson.
- Santosa, B., & Dwi, S. (2018). Work-based assessment at vocational high school in Indonesia. *International Journal of Research Studies in Education*, 8(1). <https://doi.org/10.5861/ijrse.2018.3011>
- Setneg RI. Undang-Undang No. 14 Tahun 2005 tentang Guru dan Dosen. , 19 Sekretaris Negara RI, Jakarta Indonesia § (2005).
- Sudarwanto, T., & Sulistyowati, R. (2019). *The Effect of Pedagogic Competence , Professional Competence , Social Competence , and Teachers ' Personal Competence on Teachers ' Performance in Marketing Program*. 2019, 982–987. <https://doi.org/10.18502/kss.v3i11.4065>
- Sudha, R., & Janardan, P. (2014). Item Analysis - The Steps and Interpretation. *International Journal of Comprehensive Nursing*, 1(7), 1–5.
- Tim Dikdasmen. (2018). Panduan Penilaian SMK. In *Direktorat Jenderal Pendidikan Dasar dan Manengah*. Jakarta, Indonesia.
- Wikström, C., & Eklöf, H. (2019). Educational assessment in digital environments: insights from different assessment contexts. *Education Inquiry*, 10(1), 1–3. <https://doi.org/10.1080/20004508.2019.1568779>
- Yıldız, M., & Fidan, U. (2019). The Reliability and Validity of the Fitjump Photoelectric Cell System for Determining Vertical Jump Height. *Measurement in Physical Education and Exercise Science*, 00(00), 1–9. <https://doi.org/10.1080/1091367X.2019.1673394>