


Tests of Writing in the School Examination in Upper Secondary Schools

SAGE Open
 October-December 2016: 1-7
 © The Author(s) 2016
 DOI: 10.1177/2158244016673130
 sgo.sagepub.com


Daniel Ginting¹ and Ali Saukah²

Abstract

Although the results of the final examination in Indonesia were the dominant factor in determining high school graduation, the public still did not know how the examination was administered nor how results were used to determine student graduation. Despite its importance, no comprehensive studies about its implementation have been conducted. The present study investigates the implementation of school-based assessment (SBA) in upper secondary schools focusing on the development and administration of the English writing test. This particular test was not covered in the English national examination. Twenty-one schools were surveyed, selected through stratified random sampling. In-depth case studies were conducted in three selected schools representing fully implementing school (FIS), moderately implementing school (MIS), and partially implementing school (PIS). The majority were categorized as PIS. This study suggests that the way in which examinations were implemented needs serious consideration, especially in light of the new regulation that student graduations are based on the result of the school examinations and no longer on the result of the national examination.

Keywords

school examination, writing tests, Government Regulation No. 13/2015

Introduction

The recent government regulation in Indonesia has changed the status of school examinations from being low stakes to being high stakes (Regulation No. 13/2015). In the Indonesian context, school-based assessment (SBA) in upper secondary school is intended to measure whether students have attained competence standards for graduation, and SBA can now determine whether or not students graduate. The competencies cover subject mastery, knowledge, good character, and attitude and skills necessary to become independent individuals and to continue their education.

Despite the absence of a comprehensive study of the results of Indonesian school examinations, some writers outside Indonesia have reported positive aspects of the implementation of SBA. For example, Chong (2009); Talib, Kamsah, Naim, and Latif (2014); and Tong and Adamson (2015) mentioned that SBA can promote a favorable education process that is oriented to learning. As such, schools are no longer preoccupied with teaching to test practices due to negative washback effects of the National Examination (Andrew, Fullilove, & Wong, 2002; Cheng & Watanabe, 2004; Furaidah, Saukah, & Utami, 2015; Qi, 2005). They allow teachers to get involved in making assessment decisions and improving teaching methods in

response to students' needs (Maxwell & Cumming, 2011, in Talib et al., 2014).

SBA challenges teachers' creativity in monitoring students' learning progress and designing appropriate tests aligned with the curriculum content (Sulistyo, 2009; Talib et al., 2014). In the writing test in Indonesia's National Examination, for example, teachers needed to become more creative in designing more relevant performance assessments, especially as they still use indirect testing. Indirect testing of performance reduces its validity (Cohen, 1998; Hughes, 2003; McNamara, 1996; Weir, 1993) because it is difficult to find any correlation between the micro aspects of linguistics (knowledge of grammar and vocabulary) with writing ability.

Other studies were also in doubt. In Hong Kong, for example, skepticism about the appropriateness of SBA was widespread when it was first introduced (Cheng, Andrews, &

¹Ma Chung University, Malang, Indonesia

²Universitas Negeri Malang, Indonesia

Corresponding Author:

Daniel Ginting, Ma Chung University, Villa Puncak Tidar N-1 Malang, East Java 55161, Indonesia.

Email: daniel.ginting@machung.ac.id



Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 3.0 License (<http://www.creativecommons.org/licenses/by/3.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Yu, 2011). In Malaysia, a similar issue had revolved around the technical aspects of the implementation and teacher–student readiness for change. Majid (2011) mentioned that teachers in Malaysia still had some uncertainties about the demands of SBA. In particular, the teachers were concerned about their ability to meet its demands and their role, and expected difficulties in implementation. The same findings were also reported by Talib et al. (2014), who mentioned that almost 80% of Malaysian teachers' SBA was within the range of unsatisfactory to basic. This finding implies that the teachers' knowledge of language testing was still low. Similar stories have also been reported by Sulisty (2009) when interviewing a number of Indonesian teachers long before SBA was made a high-stakes test. The teachers were not fully ready to conduct SBA and still needed expertise in language testing.

Although the prevailing laws have made all schools in Indonesia implement SBA as a criterion for student graduation, no comprehensive study has been conducted. This study seeks to fill the gap in the literature by focusing on the implementation of writing skills as one of the subjects tested in the examination at the end of upper secondary school. It begins by presenting the results of the survey in general upper secondary schools in the city of Malang (East Java Province, Indonesia), and then reports on a study carried out in three of them. The article concludes with recommendations for appropriate treatment of writing tests. These improvements contribute more widely to the effective development of SBA examinations in Indonesia.

Method

The target population of this present study included all of the general upper secondary schools in Malang conducting SBA writing tests as the examination at the end of upper secondary school. Referring to the reports of the National Examination from the Education National Standard Body Badan Standar Nasional Pendidikan (BSNP) 2009–2010, 43 general upper secondary schools in Malang were reported to have participated in the National Examination in 2010 (see Table 1).

This population reflected an imbalance; students in private and public schools had different levels of achievement. For example, the score distribution varied considerably, with the highest and lowest scores totaling 8.27 and 5.68.

For that reason, a stratified random sampling technique was used with the following procedure. First, the desired sample size was 50% of the total population (43 schools), giving 22 schools to be subjects. Next, samples were stratified on the basis of whether schools were state or private and on the basis of their achievement. This was done by establishing the quantitative categories of school achievement and the interval values for each category. The range (2.59) was obtained by subtracting the lowest score (5.68) from the highest score (8.27). Next, the interval score (0.9) for each school achievement category was established by dividing the range

Table 1. Descriptive Statistics of the 2010 National Examination of the General Upper Secondary Schools in Malang.

The average score of national examination	
<i>N</i>	43
Missing	0
<i>M</i>	7.2
Median	7.4
Mode	6.93
<i>SD</i>	0.68756
Range	2.59
Minimum	5.68
Maximum	8.27

Source: BSNP (2010).

Table 2. School Samples of the Present Study.

No.	Category	State schools	Private schools	Total
1	High-achieving schools (≥ 7.30)	5	6	11
2	Middle-achieving schools (6.49–7.29)	0	7	7
3	Low-achieving schools (5.68–6.48)	0	4	4
Total		5 (23%)	17 (77%)	22 (100%)

score of 2.59 by 3. This resulted in three categories of school: low achieving (5.68–6.48), middle achieving (6.49–7.29), and high achieving (≥ 7.30). The summary of the stratified random sampling of this present study can be seen in Table 2.

Following the survey, three different schools were studied, representing full, moderate, and partial implementation. These three schools were then categorized according to the extent to which they had implemented the writing test as a school examination, determined by a checklist of about 14 yes/no questions (a dichotomous closed question format). A *yes* answer was scored 1, a *no* answer was scored 0, and a blank answer was also scored 0. Checklist items were weighted because each question had a different degree of importance: less important (0.10), important (0.30), and very important (0.60). Of the 14 questions, two were considered as less important (14%), five were viewed as important (36%), and seven were regarded as very important (50%).

This resulted in three categories of schools for their implementation of the test (see Table 3):

Exploratory factor analysis (EFA) was carried out to assess the construct validity of the instrument (Walt & Steyn, 2008; Weigle, 2002). The qualitative data were obtained from document study and interview with semistructured format. The data analysis obtained from the checklist went through three stages. First, it dealt with the descriptive statistics. Second, this present study used the chi-square test of a distribution to obtain the evidence of the significant differences in the frequency distribution of different categories. Third, following

Table 3. Categorization of Schools Based on Writing Test Implementation.

Total score	Percentage of all checklist items	Category
0-1	45-55	PIS
2-3	64-73	MIS
≥4	82-100	FIS

Note. PIS = partially implementing school; MIS = moderately implementing school; FIS = fully implementing school.

Table 4. Cross-Tabulation Analysis of the Implementation of the Writing Tests.

Category of the implementation of the writing test	Sectors					
	State	%	Private	%	Total	%
PIS	1	20	9	56	10	48
MIS	2	40	7	44	9	42
FIS	2	40	0	0	2	10
Total	5	100	16	100	21	100

Note. PIS = partially implementing school; MIS = moderately implementing school; FIS = fully implementing school.

the chi-square computation, the data were then analyzed by means of cross-tabulation statistical technique.

Statistical Validation

The chi-square technique was used to check for significant differences among the variables under investigation: the implementation of the writing test in the state and private secondary schools. This study found that the value of the asymptotic significance (two-sided) Pearson chi-square was .024, which was smaller than the significance alpha (α) .05. As such, the approximately significant (.024) < .05 indicated that the null hypothesis (H0) was rejected. Hence, it was concluded that there was a significant difference in terms of the implementation of the writing test in the state and the private upper secondary schools.

Next, cross-tabulation was done to indicate the frequency with which the corresponding categories of the categorical variables co-occur. Referring to Table 4, the majority of the upper secondary schools (48%) were categorized as partially implementing school (PIS) category, followed by 42% as a moderately implementing school (MIS) category and 10% as fully implementing school (FIS) category.

Findings of Categorization of Writing Tests

Regarding the ratio of private to state schools, this study categorized 56% of private schools and about 20% of the state schools as partially implementing. In the MIS category, 44% were private schools and 40% were state schools. About 40%

Table 5. Preparation for Other Writing Test Materials.

Response	Frequency	%	Valid %
No	20	95.2	95.2
Yes	1	4.8	4.8
Total	21	100.0	100.0

Table 6. Consultation With the Experts.

Response	Frequency	%	Valid %
No	16	76.2	76.2
Yes	5	23.8	23.8
Total	21	100.0	100.0

of state schools belonged to the FIS category, but no private schools.

The clustering indicated that the policy on the writing tests in school examinations had not yet operated as expected; nearly all subjects were clustered in moderate and partial implementation categories.

Implementation Description

The following is a detailed description about how schools implemented the writing tests.

In the checklist, the respondents were asked whether they had made some alternative writing materials (the secondary tests and the makeup tests). Quite surprisingly, 95.2% of all schools in the sample preferred not to prepare other writing materials (see Table 5). According to the teachers in the majority of schools (MIS and PIS), their heavy teaching loads kept them from preparing alternative tests. They reported that they usually had to handle from four to six classes, representing approximately 24 teaching hours per week. Besides, they also believed that students could not cheat in the writing test because it was subjective.

The next question in the checklist asked respondents whether they had involved experts in validating assessment constructs. Table 6 shows that both MIS and PIS, as the majority (76.2%), did not consult experts either inside or outside their schools. Schools in both categories viewed the involvement of experts outside their schools as complicated because of administrative procedures. They were not sure whether the schools supported the idea of involving external experts in test development. Instead, the teachers were asked to work on their own test designs. Even if there was Musyawarah Guru Mata Pelajaran (MGMP) (a teacher association teaching in the same subjects), they never contacted them due to their heavy teaching loads. This implies that most schools did not see the involvement of experts as an urgent need.

Meanwhile, only FIS (23.8%) had involved experts in test design. These experts were usually assigned to check whether

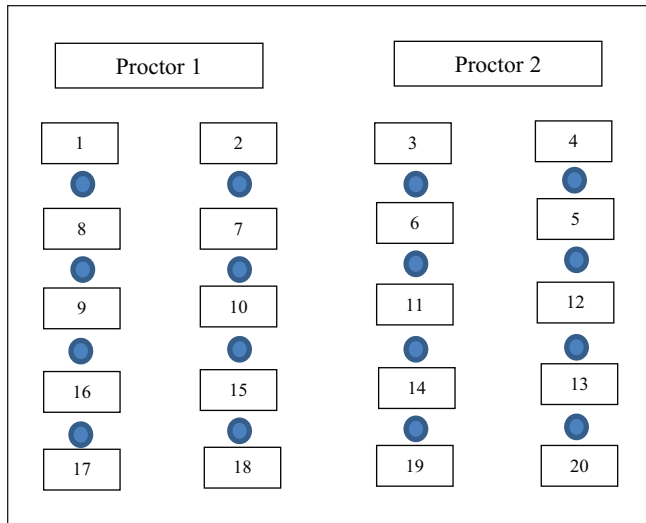


Figure 1. Seating arrangement in the school examination.

Table 7. Seating Arrangement.

Response	Frequency	%	Valid %
No	2	9.5	9.5
Yes	19	90.5	90.5
Total	21	100.0	100.0

Table 8. Sanction for the Students and Proctors.

Response	Frequency	%	Valid %
No	2	9.5	9.5
Yes	19	90.5	90.5
Total	21	100.0	100.0

the test design had fit the requirements of the construct representation and construct relevance.

The next checklist question asked respondents whether their schools had followed the regulation for seat arrangement in the test room, moving the seats of test takers 1 meter away from each other (see Figure 1). Table 7 shows that nearly all schools (FIS and MIS; 90.5%) adhered to this regulation. By contrast, PIS (9.5%) that did not follow the procedure had permitted the students to sit as in a regular class. The PIS teachers believed that fair administration of the school examination was not always primarily indicated by such a seat arrangement. Instead, they relied on thorough supervision of the teachers.

The next question asked respondents whether their schools had informed both the students and proctors about the procedures for penalties for cheating during the examination. About 90.5% of schools (FIS and MIS) reported that they had informed both students and the proctors of those procedures through oral delivery in meetings and in printed documents (see Table 8). The majority of schools (FIS and

Table 9. The Final Score.

Response	Frequency	%	Valid %
No	18	85.7	85.7
Yes	3	14.3	14.3
Total	21	100.0	100.0

MIS) adopted the rules of the test administration from the local National Education Department and disseminated them prior to the exam. Moreover, teachers in FIS and MIS reported that the schools posted the rules on the walls of all testing rooms so that everyone could see them.

By contrast, PIS (9.5%) reported the rule was only disseminated orally prior to the examination. Teachers in PIS mentioned that it was not necessary to inform proctors of penalty procedures; violating the norms of the test administration would never make any sense to proctors as they were good people tied to ethical codes.

The next item in the checklist asked respondents whether they as assessors had shared their scores to obtain the average as the final score. While MIS and PIS, as the majority (85.7%), never carried out this procedure, FIS (14.3%) had used this strategy in the scoring procedure (see Table 9). FIS headmasters assigned a teacher to carry out cross-scoring. All assessors took turns to read all students' work independently and then entered the scores in the forms. The students' final scores were obtained by dividing all scores. This achieved greater objectivity in scoring.

MIS and PIS, the majority of schools (85.7%), did not have a cross-scoring procedure and never applied average scores to obtain final scores. For them, cross-scoring was time-consuming and was impossible due to the limited number of assessors (three to five teachers). Some reported that more than 400 students took the tests. For practicality, teachers decided to divide students' writings into equal piles. For example, if there were 100 items and two assessors, then, each assessor would get 50 items to read and score. They only read and scored students' writings of their own, and never shared those pieces with other assessors to read and score. The school administrators often demanded that they had to submit the scores in short time frames, with some teachers even reporting that they were required to finish only 1 day after the examination.

As another alternative, MIS conducted a cross-prompt assessment procedure. Different teachers were assigned particular prompts in the test. For example, if there were four writing prompts in the writing test, teachers would only read and score the prompts to which they had been assigned. As Teacher A had been assigned to read Prompt 1, he or she only scored that part and left other prompts to other teachers. In the same way, Teacher B was assigned to read Prompt 2, and he or she only checked and scored that part. The same procedure was also true with Teachers C and D. Hence, the final score was obtained by adding all scores from Teachers A, B, C, and D. These teachers believed that this procedure potentially increased the fairness of the scoring. The assumption was that

Table 10. The Involvement of the Third Rater.

Response	Frequency	%	Valid %
No	19	90.5	90.5
Yes	2	9.5	9.5
Total	21	100.0	100.0

Table 11. The Scoring Rubrics.

Response	Frequency	%	Valid %
No	18	85.7	85.7
Yes	3	14.3	14.3
Total	21	100.0	100.0

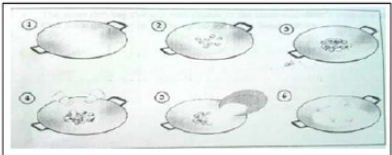
teachers easily tend to give unfairly higher scores to their own students.

The next item on the checklist was to ask respondents whether they had involved third assessors to resolve discrepancies in scores. MIS and PIS, as the majority (90.5%), never involved a third assessor to mediate score discrepancies (see Table 10). For them, this procedure was unnecessary and they preferred to score the students' work individually. In addition to the limited number of assessors, scoring the students' work individually was more practical as they could submit the score list in due time.

Meanwhile, FIS, as the minority (9.5%), that had used a cross-scoring procedure viewed the involvement of the third assessors as necessary to solve score discrepancy problems. According to FIS teachers, besides mediating the score discrepancy, the third assessors were also assigned to check the assessors' lists of scores. When discrepancies inevitably occurred, the third assessor usually asked all assessors to meet to discuss them. If they came to agreement when reexamining their previous grades, the two assessors simply modified the scores. However, if they could not get agreement, then the third assessor was assigned to read the students' work and give their own scores. In these cases, students' final scores were the mean of the scores from all assessors, including the third assessors.

Another item in the checklist asked whether respondents had prepared scoring rubrics to score the students' essay (see Table 11). In practice, MIS and PIS as the majority (85.7%) had used mixed scoring formats both for essays and objective tests. For objective tests, they prepared the answer keys to such test prompts: filling in the empty blanks with the correct words and arranging the jumbled sentences into good paragraphs (see Figures 2 and 3). Meanwhile, for essays, the scoring rubrics were designed in an analytical format that included such aspects as grammar, mechanics, vocabulary, and content. Hence, the students' final scores were the combination of the results from both objective tests and essays.

The following pictures tell us how to make fried rice. Rearrange this jumbled text. The picture will help you to arrange it.



- Here is the process of making fried rice
- Slice the shallots and garlic. Fry the shallots and garlic until it smells nice.
- Pour the cooking oil into the pan and heat it.
- Fried rice is very common in our society. Many people like to eat it. Making fried rice is not a difficult thing. We just need a plate of rice, cooking oil, tomato sauce, pepper, shallots and garlic.
- Stir the rice on a plate. Sprinkle fried crispy onion. Serve it while it is hot. You can eat it with an omelette, fried chicken and some vegetables, such as cucumber and lettuce.
- Put the rice into the pan.
- Season with some salt, pepper and tomato sauce.

Figure 2. The sample of objective writing prompts in MIS.

Note. MIS = moderately implementing school.

Arrange the jumbled sentences below into a good paragraph

CHEESE OMELET

1. Then, whisk the eggs with a fork until smooth
2. Add some milk and whisk well
3. Heat the oil in a frying pan
4. Crack the eggs into a bowl and stir
5. Grate the cheese into the bowl and stir. Then
6. Pour the mixture into the frying pan
7. Turn the mixture with a spatula when it browns. Cook both sides
8. The cheese omelet is ready to be served
9. Place it on a plate and season it with salt and pepper
10. After the omelet is cooked

Figure 3. The sample of objective writing prompts in PIS.

Note. PIS = partially implementing school.

Meanwhile, FIS, as the minority (14.3%), had used a holistic scoring rubric. Some descriptors even described certain levels of competence for each score (5, 10, 15, 20, and 25). FIS also trialed their writing tests prior to implementation as the real test (see Figure 4). A small number of the students were invited to take the test, and the results were gathered for analysis. One teacher with extensive experience in language testing took roles of both expert and chairperson of the teacher panel that was responsible for validating the test.

The next question asked respondents whether they had trialed the writing test before using it. This step collected evidence on whether the particular test models and scoring rubrics had been properly designed. Table 12 shows MIS and PIS, the majority of schools (95.2%), did not trial the test. Instead, the teacher panel discussed the test quality. PIS that had a shortage of teaching staff entrusted the teacher to assess the quality of their own tests.

By contrast, the teachers in FIS (4.8%) stated that the trial was part of a series of activities in test development. Teachers observed the test administration and analyzed the results of the writing tests. During the trial, about three or four voluntary students did the writing tests. The teachers gathered the responses of the sample students and reviewed them. They

WRITING TEST			
SCHOOL YEAR 2013/2014			
Subject	: English	Day/Date	: Thursday, December 12, 2013
Class	: XII Language	Time	: 120 minutes
Do the writing test clearly!			
1. Create an Hortatory Exposition Text with a theme of “Clean and green is our School”			
(Persuade the reader to support the school program to get the environment of the school clean and green) Start your writing with:			
	Thesis	: An introduction of the issue	
	Arguments	: Arguments to lead the readers believe	
	Recommendation	: Strengthen the writer’s point of view to ask the reader to follow	
Use	: Simple present tense and passive voice		
	Connectors , examples: then, after that, etc.		
	Words that link arguments, examples: firstly, on the other hand, therefore, etc.		

Figure 4. The sample of writing prompts in FIS.

Note. FIS = fully implementing school.

Table 12. The Tryout of the Tests.

Response	Frequency	%	Valid %
No	20	95.2	95.2
Yes	1	4.8	4.8
Total	21	100.0	100.0

Table 13. The Scoring Under Controlled Reading.

Response	Frequency	%	Valid %
No	18	85.7	85.7
Yes	3	14.3	14.3
Total	21	100.0	100.0

focused on noting areas of students’ confusion, vague or incomplete responses, and unanticipated responses. If students’ responses indicated ineffectiveness, considerable restructuring of the testing tool would be necessary.

The next question in the checklist asked respondents whether they had scored the students’ writing at the same time and place in a group work. Table 13 shows that MIS and PIS, the majority of schools (85.7%), had disregarded this procedure in the writing test administration. They did not see such a procedure as important. To them, this requirement was impractical due to possible distractions from other members of the group and the limited number of assessors, so they preferred to score students’ work individually. Thus, the teachers shared the students’ writing in equal number and took them home for further scoring.

By contrast, FIS regarded the controlled reading procedure and scoring as a necessity. The headmasters in FIS (14.3%) officially required that the reading of the students’ writing took place at the school within the scheduled time. It was done at a certain date and place and led by one teacher

who had been appointed as convenor. The convenor usually announced the technical aspects of the scoring procedure with which assessors had to comply, such as reading the students’ writing attentively, applying the scoring rubrics to the students’ writing, using the students’ writing samples to represent different levels of performance or the aspects being assessed, putting the students’ score on the list of the grades provided, obtaining the final scores, and involving the third assessor in case of a discrepancy in scores.

Discussion

This study found the writing test had not been satisfactorily implemented. In practice, nearly all schools were clustered into PIS (48%) and MIS (42%), with very few categorized as FIS (10%). Many schools neglected substantial aspects of language testing, such as validity and reliability (Bachman & Palmer, 1996; Brown, 2007; Hamp-Lyons, 1991; Weigle, 2002). For example, the threat to validity is obviously apparent for most schools (MIS and PIS), which still used indirect testing for assessing writing skills. Besides the absence of experts and test trialing, many teachers still believed that an indirect testing approach, as used in the National Exam, is credible. In line with Sulisty’s (2009) findings, most teachers readily adopt indirect testing to develop writing tests in the SBA because none of them doubted the validity of the National Examination. Moreover, the threat to reliability is quite observable; most schools never scored under controlled reading, never used cross-scoring, and never involved a third assessor to resolve score discrepancies between assessors.

Although the School Examination test has high stakes for students, it did not affect teachers’ professional performance. The present study has shown how teachers were reluctant to be creative in developing a more relevant assessment to test performance skills like writing. Saukah and Cahyono (2015) have reported that teachers in mostly low-achieving schools

have given their attention to the preparation of the National Examination, although it is no longer the sole basis for student graduations. They mention heavy workloads and the absence of institutional support as reasons for not developing their professional skills. These findings were similar to the study by Talib et al. (2014). Teachers were unprepared for the change and found the new system challenging. About 79.66% of Malaysian teachers were not fully engaged in practicing SBA. Similarly, overall results revealed that Malaysian teachers' SBA practice classrooms are within the range of unsatisfactory to basic levels of almost 80%.

Conclusion

This study finds that the practices of the writing test in the final examination of secondary school have resulted in different patterns of implementation, with a strong tendency for the partial implementation of policy. Hence, this study recommends some action points. First, teachers should be aware that the school examination, as a high-stakes test, demands significant responsibility on the part of teachers, and this should affect their professional performance. The decisions that teachers make in the school examination will affect the future of their students. Second, although the government has so far provided the teachers with technical guidance for administering the examinations (*Pedoman Teknis Pelaksanaan Ujian Sekolah/Madrasah/The Technical Guidance of the Implementation of Madrasah/School Examination*), teachers clearly need professional development to improve their skills. Short training courses given by experts would help teachers to be able to design better tests for school examinations.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

- Andrew, S. J., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System, 30*, 207-233.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Brown, H. D. (2007). *Principles of language learning and teaching*. New York, NY: Pearson Education.
- BSNP (2010). *Laporan Hasil Ujian Nasional Tahun Pelajaran 2009-2010* (Computer software). Jakarta: Pusat Penelitian Pendidikan Badan Penelitian dan Pengembangan Kementerian Pendidikan Nasional.
- Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing, 28*, 221-249.
- Cheng, L., & Watanabe, Y. (2004). *Washback in language testing: Research contexts and methods*. Thousand Oaks, CA: SAGE.
- Chong, K. K. (2009). Whither school-based coursework assessment in Singapore. In *35th IAEA Conference-Assessment for a Creative World*, Brisbane, Australia. Retrieved from http://www.iaea.info/documents/paper_4d73afd.pdf
- Cohen, A. (1998). *Strategies in learning and using a second language*. Harlow, UK: Longman.
- Furaidah, F., Saukah, A., & Widiati, U. (2015). Washback of English national examination in the Indonesian context. *TEFLIN Journal, 26*(1), 36-58.
- Hamp-Lyons, L. (1991). *Assessing second tests*. Harlow, UK: Pearson Education.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Majid, F. A. (2011). School-based assessment in Malaysian schools: The concerns of the English teachers. *Journal of US-China Education Review, 8*(10), 1-15.
- Maxwell, G. S., & Cumming, J. J. (2011). Managing without examination: Successful and sustained curriculum and assessment reform in Queensland. *Australia's Curriculum Dilemmas: State Perspectives and Changing Times, 202-222*.
- McNamara, T. (1996). *Measuring second language performance*. London, England: Longman.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*, 142-173.
- Saukah, A., & Cahyono, A. E. (2015). National exam in Indonesia and its implications to the teaching and learning of English. *Jurnal Penelitian dan Evaluasi Pendidikan, 19*, 243-255.
- Sulistyo, G. H. (2009). English as a measurement standard in the national examination: Some grassroots' voice. *TEFLIN Journal, 20*(1), 1-24.
- Talib, R., Kamsah, M. K., Naim, H. A., & Latif, A. A. (2014). From principle to practice: Assessment for learning in Malaysian school-based assessment classroom. *International Journal of Social Sciences and Education, 4*, 850-857.
- Tong, S. A., & Adamson, B. (2015). Student voices in school-based assessment. *Australian Journal of Teacher Education, 40*(2), Article 2. doi:10.14221/ajte.2015v40n2.2
- Walt, J. L., & Steyn, F. (2008). The validation of language tests. *Stellenbosch Papers in Linguistics, 38*, 191-204.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.
- Weir, C. (1993). *Understanding and developing language tests*. New York, NY: Prentice Hall.

Author Biographies

Daniel Ginting is currently teaching at the English Letter Study Program, Ma Chung University (Indonesia). He obtained his Doctor's degree in ELT from State University of Malang.

Ali Saukah has been a full-time faculty member of the English Department of Universitas Negeri Malang (Indonesia) since 1980, and involved in the development of English Language Curriculum and in other related activities at the national levels since 1990.