

# Designing Bilingual Test Items: *Teacher's Strategies*

Euis Yuyun, Yanty Wirza  
Universitas Pendidikan Indonesia  
Bandung, Indonesia  
yuyunfillah@upi.edu

**Abstract**—The study is aimed at exploring teacher's strategies in designing bilingual test items in science and mathematics. Through employing a qualitative approach and using an interview to collect the data, the study involves an elementary teacher in a private school that implements a bilingual program in the classroom. The study reveals several findings: 1) the teacher uses a simplified language for the items; 2) given students' barrier in terms of understanding specific terms, the teacher aids the students by using L1 to ensure students' understanding; 3) because of students' limited English proficiency, the teacher cautiously chooses the familiar words in the items to be translated into English; 4) the teacher seeks help to ensure the test items are understandable.

**Keywords:** *bilingual testing, content assessment, elementary school, teacher strategy*

## I. INTRODUCTION

Bilingual Education (BE) is defined as a dual-focused learning approach in which subject matters are taught through foreign languages or the languages that are not the dominant/first languages (Coyle, Hood, & Marsh, 2010; Dalton-puffer, Llinares, Lorenzo, & Nikula, 2014). The related terms are also used to represent the same programs such as EMI (English-Medium of Instruction) that are commonly addressed to the context of Anglophone countries in which English is as a foreign language, Content and Language Integrated Learning (CLIL) that reside in Europe, Content-Based Instruction (CBI), English for Specific Purposes (ESP), Bilingual Education and Content-based language teaching that are more familiar in American contexts (Hamid, Thi, Nguyen, & Baldauf, 2013; Lasagabaster & Sierra, 2010; Wannagat, 2008).

The mushrooming of bilingual schools in Indonesia was initially triggered by the emergence of international school namely TIS (True International School). Hamied (2013) asserts that the issue of globalization and being international has significantly influenced all aspects including Education leading the nation to commit with English. In Indonesia, the establishment of ISS (International Standard School) or also called *Rintisan Sekolah Bertaraf Internasional* (RSBI) had reflected the issue of acquiring international status. Thus, the demand for globalization urged the Indonesian government to establish and expand at a minimum one international school in each district by 2014 (Zacharias, 2013). However, in practice, ISS implementation came out to be problematic since it widens the social pressures. For example, the assumption of being elite or coming from a wealthy family is attached to the students who studied there. The middle or lower status of a family, on the other hand, hardly afford to study there as its expense is too high (Hamid, Jahan, & Islam, 2013). Even though ISS policy

had been terminated in 2013, there are several private schools (national plus) running bilingual programs that combine national curriculum with international curricula such as Cambridge or International Baccalaureate (IB). These schools use English as the medium of instruction in the classroom to teach subject matter such as mathematics and science or other selected subjects. Zacharias (2013) believes that the practice of EMI (English-Medium of Instruction) policy in Indonesia continues to be pervasive.

The studies related to content assessment have been massively conducted in non-English speaking countries, for instance, in Pennsylvania (Clark-Gareca, 2016; Kachchaf et al., 2016; Lee, 2018; Llosa et al., 2016). Most of these studies highlight the topics as to test accommodation, English language proficiency standard, and content standard, linguistic modification, linguistic complexity. Besides, the studies are inclined to focus on the context of Bilingual Education for English Language Learners (ELLs) in which English as the first language. Referring to the aforementioned studies, it is rarely found the studies dealing with the teacher's strategies on composing bilingual test items at bilingual education in the context of non-English speaking countries, moreover, most of the studies related to Bilingual Education in non-Anglophone contexts have been limited to the implementation of EMI policy (Bilingual Education policy); India (Endow, 2018), China (Wei & Feng, 2015), Kenya (Hsieh, Lonescu, & Ho, 2017) Pakistan (Rashid, Muzaffar, Dar, & Butt, 2016). Thus, this study aims to explore teacher's strategies in designing bilingual test items (English and Indonesian) for assessing the students' performance in a bilingual classroom wherein mathematics and science are taught in English.

## II. METHOD

### A. Research Design

The article entitled teacher's strategies of designing bilingual test items adopts a case-study approach. A case study is conducted to deeply understand complex social phenomena (Yin, n.d.), in other words, a case study is intended to scrutinize a phenomenon holistically and meaningfully. The study investigates one phenomenon in education that is how the teacher's strategies on constructing test items in a bilingual classroom in which English is used as a medium of instruction in teaching subject matter mathematics and science.

### B. Participant and Setting

The study is to seek teacher's strategies on designing bilingual (Indonesian-English) test items in an elementary school that enacts the bilingual program in which mathematics and science are taught in English. The textbooks of math and science used by the teacher in the classroom are bilingual; English and Indonesian. Meanwhile, the content areas are

designed based on the national curriculum, K-13 (Curriculum 2013). The school offers two programs; bilingual and non-bilingual programs. Non-bilingual (regular) classroom uses Bahasa Indonesia as a medium of instruction, meanwhile, bilingual classroom uses English as a medium of instruction including in teaching mathematics and science.

The participant of this study is one teacher at a bilingual classroom. This participating teacher is a homeroom teacher of fifth grade who has been teaching there since 2015. She got a bachelor's degree in English literature and has been involved in designing bilingual test items of both math and science for eight times. She teaches English, mathematics, and science in her classroom (science and mathematics are mandated to be taught by the homeroom teacher). The participant is purposefully selected seen from her experience of getting involved in designing content assessment for science and mathematics. The purposefully selected participant aids the researcher understand the problem being investigated (Creswell, 2014).

#### *C. Data Collection Procedures*

A face-to-face semi-structured interview was held with the participating teacher. The researcher beforehand prepared a list of interview questions but in practice, it was conducted flexibly. A list of questions is used as a reminder that will cover the main topics. The interview lasted for about 30 minutes and it was digitally recorded using sound recorder on the mobile phone. The questions were delivered in Bahasa Indonesia but the interviewee sometimes code-switched and answered the questions in English.

#### *D. Data Analysis*

The data related to the teacher's strategies in designing content assessment gained from the interview was analyzed using qualitative content analysis (Miles, Huberman, & Saldana, 2014). The steps taken include; familiarization (transcribing the data), sorting and coding the data, identifying themes and patterns, developing ideas, making interpretation and conclusion. The process of categorization was conducted based on Kopriva's (2008) framework of constructing test items for ELL; Participation in content test development, test specification, content accessibility and validity, and also testing accommodation.

### III. FINDINGS AND DISCUSSION

#### *A. Content Expert Participation*

Participation in test development deals with the involvement of the connected elements during developing the test. The students need to be involved during the stages of test development and field test to ensure that the system of assessment represents the students' attainment and equivalent accuracy (Kopriva, 2008). It includes expert participation and student participation either in small pilots or large field tests.

Based on the data obtained through the interview, it was found that there is expert participation in designing test items for math and science assessment.

Because there are three branches of the school, the task to design the test items is handled by these three schools. Last year, it was the third branch's turn and this year, it is our

school's turn. In our school, the test maker for science is handled by Ms. R because her major is science education but she only constructs the items in Bahasa Indonesia, for the English version is assigned to each homeroom teacher of the bilingual class. For math tests, this year it is the third school's turn.

In designing math and science test, the teacher receives the assessment in Bahasa Indonesia which is then translated into English. This is done since the materials of science and math in regular classes (not bilingual) are more comprehensive rather than the materials in the bilingual classroom. Regular classes adopt a local textbook that provides complete materials for teaching, meanwhile, bilingual classrooms adopt a bilingual textbook (English-Bahasa Indonesia) but it cannot cover all the materials. Therefore, at the onset, science, and math test is designed in Bahasa Indonesia since it is a teacher from regular class who understands coverage of the materials and is an expert of these subject matters. It was found that the test maker is the teacher who is an expert at the subjects seen from the major of the educational background.

In terms of students' participation, the teacher who translates the assessment used to have a kind of pre-test to her classroom to identify students' level of understanding of science and math. However, the assessment did not cover all the test specifications; rather, it solely focused on fill-in-blank questions without multiple choices and essays.

I have tried provided the students with a pre-test before the test is given to them but for math only and I only give the fill-in-blank questions. Even though it is not completed by multiple choices and essays, the pre-test still covers the materials already learned in the classroom.

Even though only employing one test specification, the test could cover the coverage of materials. Therefore, the teacher used students' participation (pre-test) as a consideration in designing the test items even though it is only a small sample pilot rather than a large sample.

#### *B. Test Specifications*

Test specifications deal with the outline of testing, a kind of blueprint as to the coverage of testing (Kopriva, 2000). Test specifications are reflected from content standards or the descriptions of subject matter to what items are going to be included in the test. The components of the test specifications comprise *breadth* covering knowledge and skills and *depth* covering the complexity of the test items.

In terms of breadth, it is not the teacher who determines the knowledge and skills to be presented in test items; rather, it is designed by the teacher of science and math from regular class (non-bilingual). The vice-principal clarifies that before this policy, the school mandated each homeroom teacher of the bilingual classroom to design test items bilingually regardless of their content expertise. However, it was found a significant gap in content coverage of test items. The bilingual test items could not cover all the materials included in the curriculum differing from that of test coverage of regular class, thus, designing test items is centralized at one school.

The items are centrally designed in Bahasa Indonesia since the test is also made for a regular class, it is then translated into English by bilingual teachers.

As to the content coverage between bilingual and bilingual textbook is also elaborated by the teacher;

Science and math textbook is different from that of textbooks used by regular class; in terms of content, textbooks for a regular class are more comprehensive. But thankfully we have a worksheet, so, it helps me.

In terms of depth, the teacher herself who decides the items to be translated considered from the complexity of the words.

I choose which items will be translated into English. I consider the word complexity. I usually look for the common words which are already understood by my student. I also consider the length of the sentence in the item; I do not convert the item to English if the question is too long.

She also explains that the portion is about 80 % English and 20 % Bahasa Indonesia comprising multiple choices, essays, and short answer questions. She chooses an item consisting of common words. This method is chosen since the class she handles is a lower level class in which her students' language proficiency is limited. Additionally, the numbers of items are not put into concern as long as she finds simple words in an item. In other words, the translation is conducted randomly. Word choice is part of her constraints in translating test items into English since she has to match the translation with students' capabilities.

In terms of content accessibility and testing accommodation, the teacher elaborates the following;

*She uses simplified language for the items*

When the teacher asked how she chooses the items to be translated,

It depends on the items, if the item is too long, just make it short. There are words cannot be translated but without the word, the sentence can still have meaning

The teacher tried to simplify the language for the sake of students' comprehension of the items. The main consideration is how the item can be understood by the students since she realizes that her classroom is categorized as a lower achievement with LEP, she tries to simplify the language of the test items. This is in line with Young et al., (2014) that the teacher might develop a linguistic modification to ensure the test is accessible for the students. One of the modifications is simplified vocabulary, that is, the identification of challenging words which are not related to the content and replace the words with accessible ones.

*Because of her students' limited English proficiency, she cautiously chooses the items to be translated through selecting the commonest words*

Furthermore, she elaborates on her constraint of doing translation in test items which leads her to provide testing accommodation for all students.

There must be the difficulty, how to make students understand the question. There are times when we do not understand particular words and then looked up in the dictionaries, the words are too difficult for the students.

The teacher extremely considers students' capacity in language mastery. She realizes that her class is dominated by students with limited English proficiency that she carefully chooses the test items to be translated into English. This is also seen from her statement;

So, language is the main constraint, it might due to students' lack of language practice which hampers them to understand complex questions and so, I have to find out appropriate words on the test. Once I found a material that has not been delivered to the students, I did not translate it.

Here she also meticulously analyses the items, if she finds the item with the material not been delivered in the classroom, she did not convert the item to English.

*Given students' barrier is in terms of understanding specific terms, she aids the students through using L1 language*

It is found that the teacher count on students' L1 in explaining a complicated item, that is, understanding particular terms related to science and math. She asserts that

When I have to translate parts of the body, hipbone, I translated into English even though I had to tell its meaning to my students; I have taught it to my students many times but they still do not understand.

There are particular terms which are still perplexing to the students that they cannot comprehend the English version of the item. Thus, the teacher needs to explain the meaning in Bahasa Indonesia. However, such a condition solely occurs in the content assessment. It is scarcely found students encounter a problem during English assessment;

Students rarely ask many questions during the English test because the coverage of English materials is taken from students' daily life.

It showed that students' difficulty is in terms of understanding specific terms in science and mathematics since the materials tend to be complex rather than English. She also asserts that whatever the students' level is, they have to pass through the test even though they might end up with remedial testing. It is unfeasible for her and the school to adjust test items with another class.

The teacher, in this case, used the students' L1 as a pedagogical function. Concerning the use of L1 in learning subject matters, Cahyani, Courcy, & Barnett (2016) reveals the function of code-switching in the classroom to construct knowledge; about this function the teacher switches between languages to ensure students' understanding as to certain topic; for instance, to understand the topic being discussed, the teacher code switched the certain word to students' L1. What the teacher did was for the sake of students' understanding. This might occur when dealing with a complex concept, difficult mathematical terms, or ambiguous words, and students' limited English proficiency (Naidoo, 2016).

*She seeks to help herself to ensure the test items are understandable*

Another problem she encountered as translating test items is arranging well-structured sentences.

Grammar and structure make me difficult to translate. Sometimes I ask for the correct translation since I know my grammar isn't good yet.

Then when she was asked whether she looks for assistance in translating she answered;

I always search in Google, Google does not always make good sentences when we translate, so, if the translation is still incorrect and I cannot fix it, I consult with my partner.

To ensure the correctness of her translation, she counts on Google translate and also peers. However, comparing to translating math test items, she asserts that it is easier to translate math assessment than science. She explains that math tends to be easier since skills and knowledge are about *counting*.

The sentence also tends to be so simple, for instance, *the result of ... is ...* that makes the students quickly accomplish the test. The essay form can also be easily understood by the students since the story conveyed on the item links to daily life. Compared to science, the terms are more difficult to understand.

Therefore, the students' barriers are mostly in terms of understanding specific terms. In this case, she sees that math is easier than science in terms of constructing the item even though the terms of content remain to be difficult for the students to understand.

The problem encountered by the teacher also comes from the students with limited understanding of addition, subtraction, multiplication, and division. She believes that if this core knowledge of math is not mastered, it will be difficult for them to attain advanced materials such as fractions, and so forth. Thus, the students with low achievement in content areas encounter difficulty in understanding items presented in the test.

#### IV. CONCLUSION

The study focusing on teacher's strategies on designing bilingual test items has revealed several salient findings; the content test takers are limited English proficient since it was said by the teacher that her class is categorized as non-optimal bilingual classroom dominated by under-achievers. Her concern is in terms of accommodating students to ensure they understand the test items and specific terms related to science and mathematics. In other words, the teacher should accommodate content assessment for students with a limited vocabulary. Due to this condition, word choice becomes her main concern in converting test items into English. Selecting particular words in converting test items into English means that the teacher is aware that there are several words (basic words) which cannot be understood since it conveys the meaning of the sentence (August, Carlo, Dressler, & Snow, 2005), therefore, replacing the complex words with the simple ones remain as the strategy employed by the teacher. It also implies that the teacher has minimized construct-irrelevant variance to provide accessibility of test items (R. J. Kopriva, 2008; Young, 2009). Apart from

word choices, the teacher also takes advantage of students' first language instead of merely focusing on the target language. As she found the students could not catch the meaning of test items, she ended up telling the keywords in Bahasa Indonesia, it typically deals with particular terms related to science and mathematics. Even though she perceives that designing mathematics test items is easier than science, the unfamiliar terms remain difficult for students. Therefore, the key consideration of constructing English test items is in terms of word complexity rather than the selection of the items based on the order of its numbers, test specifications such as multiple-choice, short answer, or essays. In this case, the teacher extremely takes the students' capability into account.

Students' level of achievement also becomes a quandary. On one side the teacher frequently delivers content of science and mathematics in the classroom including explaining the terms in math and science, however, the students remain puzzled with the words. The students are supposed to have understood all she has taught but her class level is different from that of optimal class. Although her class level is categorized as lower, the test items cannot be designed differently based on their level. On another side, delivering test items in English is also school mandatory which cannot be taken for granted. About this phenomenon, the researcher found teacher's strong commitment and perseverance as a pivotal role, how she is consistent to concurrently implement school's policy as to bilingual program and accommodates students' capability for participating in content assessment even she tried hard to present proper English test items through using Google translate and peers as her guidance.

To sum up, the teacher positively perceives content test is presented bilingually (English and Bahasa Indonesia) amid her challenges in terms of students' capability in English, students' content achievement and her (the teacher) limited English ability as well. The related studies for testing accommodation would be the recommendation. It is suggested that the students are not classified into different classes and levels based on their achievement and proficiency. If so, the school has to provide them with fair assessment and accountability (accommodation for all students) enabling students to demonstrate their knowledge and skills of subject matter (Abedi, Hofstetter, & Lord, 2004; Kachchaf et al., 2016). Besides, the school might require English Language Proficiency Standard (ELP) and content achievement standard for the students as pre-requirements to enter a bilingual class.

#### REFERENCES

- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1–28. doi: 10.3102/00346543074001001
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*(1), 50–57. doi: 10.1111/j.1540-5826.2005.00120.x
- Cahyani, H., Courcy, M. De, & Barnett, J. (2016). Teachers' code-switching in bilingual classrooms: Exploring pedagogical and sociocultural functions. *International Journal of Bilingual Education and Bilingualism, 21*(4), 1–16. doi: 10.1080/13670050.2016.1189509
- Clark-Gareca, B. (2016). Classroom assessment and English language learners: Teachers' accommodations implementation on routine math and science

- tests. *Teaching and Teacher Education*, 54, 139–148. doi: 10.1016/j.tate.2015.11.003
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning* (1<sup>st</sup> Edn.). Cambridge: Cambridge University Press.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4<sup>th</sup> Edn.). California: Sage.
- Dalton-puffer, C., Llinares, A. N. A., Lorenzo, F., & Nikula, T. (2014). "You can stand under my umbrella": Immersion, CLIL and bilingual education . A response to Cenoz, Genesee, & Gorter (2013). *Applied Linguistics*, 35(2), 213–218. doi: 10.1093/applin/amu010
- Endow, T. (2018). Inferior outcomes: Learning in low-cost English-medium private schools—a survey in Delhi and national capital region. *Indian Journal of Human Development*, 12(1), 1–19. doi: 10.1177/0973703018779725
- Hamid, M. O., Jahan, I., & Islam, M. M. (2013). Medium of instruction policies and language practices, ideologies and institutional divides: Voices of teachers and students in a private university in. *Current Issues in Language Planning*, 14(1), 37–41. doi: 10.1080/14664208.2013.771417
- Hamid, M. O., Thi, H., Nguyen, M., & Baldauf, R. B. (2013). Medium of instruction in Asia: Context, processes and outcomes. *Current Issues in Language Planning*, 14(1), 37–41. doi: 10.1080/14664208.2013.792130
- Hamied, F. A. (2013). ELT intricacies within the Indonesian language policy. In Bigalke, T. W., & Sharbawi, S. (Eds.), *English for ASEAN Integration: Policies and Practices in the Region*. Bandar Seri Begawan: IELTS, pp. 32–40.
- Hsieh, C.-N., Lonescu, M., & Ho, T.-H. (2017). Out of many, one: Challenges in teaching multilingual Kenyan primary students in English. *Language, Culture and Curriculum*, 31(2), 199–213. doi: 10.1080/07908318.2017.1378670
- Kachchaf, R., Noble, T., Rosebery, A., O'Connor, C., Warren, B., & Wang, Y. (2016). A closer look at linguistic complexity: Pinpointing individual linguistic features of science multiple-choice items associated with English language learner performance. *Bilingual Research Journal*, 39(2), 152–166. doi: 10.1080/15235882.2016.1169455
- Kopriva, R. J. (2000). *Ensuring accuracy in testing for English language learners* (Washington). The Council of Chief State School Officers.
- Kopriva, R. J. (2008). *Improving testing for English language learners*. New York: Routledge.
- Lasagabaster, D., & Sierra, J. M. (2010). Immersion and CLIL in English: More differences than similarities. *ELT Journal Volume*, 64(4), 367–375. doi: 10.1093/elt/ccp082
- Lee, O. (2018). English language proficiency standards aligned with content standards. *Educational Researcher*, 20(10), 1–11. doi: 10.3102/0013189X18763775
- Llosa, L., Lee, O., Jiang, F., Haas, A., O'Connor, C., Van Booven, C. D., & Kieffer, M. J. (2016). Impact of a large-scale science intervention focused on English language learners. *American Educational Research Journal*, 20(10), 1–30. doi: 10.3102/0002831216637348
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis* (3rd Edn.). SAGE Publications.
- Naidoo, J. (2016). Language matters! Exploring promise's use of pedagogic strategies in her mathematics classroom. *Language Matters*, 47(3), 372–392. doi: 10.1080/10228195.2016.1233572
- Rashid, A., Muzaffar, I., Dar, F., & Butt, S. (2016). The issue of English as a medium of instruction in primary schools in Pakistan: Learning English, mathematics or science? *Human Rights in Language and STEM Education: Science, Technology, Engineering and Mathematics*, 179–204. doi: 10.1007/978-94-6300-405-3\_11
- Wannagat, U. (2008). Learning through I2-content and language integrated learning (CLIL) and English as medium of instruction. *International Journal of Bilingual Education and Bilingualism*, 10(5), 37–41. doi: 10.2167/beb465.0
- Wei, R., & Feng, J. (2015). Implementing CLIL for young learners in an EFL context beyond Europe. *English Today*, 31(1), 55–60. doi: 10.1017/S0266078414000558
- Yin, R. K. (n.d.). Case study research: Design and methods. In *Applied Social Research Methods Series* (2<sup>nd</sup> Edn., Vol. 5). London: SAGE Publications.
- Young, J. W. (2009). A framework for test validity research on content assessments taken by English language learners. *Educational Assessment*, 14(3–4), 122–138. doi: 10.1080/10627190903422856
- Young, J. W., King, T. C., Hauck, M. C., Ginsburgh, M., Kotloff, L., Cabrera, J., & Cavalie, C. (2014). Improving content assessment for English language learners: Studies of the linguistic modification of test items. In J. Carlson (Ed.), *ETS Research Report Series*. doi: 10.1002/ets2.12023
- Zacharias, N. T. (2013). Navigating through the English- medium-of-instruction policy: Voices from the field. *Current Issues in Language Planning*, 14(1), 37–41. doi: 10.1080/14664208.2013.782797