

Considerations in developing and using CAT to assess students' proficiency in English as a foreign language in Indonesia

Ignatius Harjanto
Davy Budiono

Abstract. *The English curriculum for Secondary Schools in Indonesia requires the students to acquire ability in accomplishing tasks adequately, to find solutions, and to realize them in real situations. Such objectives can be achieved if, among others, teachers are qualified. This means that the graduates of teacher training colleges (TTCs) must be competent in the field having knowledge and skills. In case of English education, they must be proficient in the four English skills. To ensure their English proficiency, a test administered institutionally, which varies from one college to another, is not enough. A national test on English proficiency to the graduating students of TTCs is needed so that a national standard of proficiency for teachers can be measured. However, administering a national English proficiency test is expensive and cumbersome as well. A solution to the problem is the application of computer adaptive tests (CAT). This article argues that CAT enables the students to measure their own level of English proficiency inexpensively. Ideas on the rationale, application, and impacts of standardized proficiency test using CAT on students with learning English difficulties in regular schools are shared.*

Key words: *assessment, computer adaptive test, English proficiency, students with learning difficulties, teacher training colleges*

Introduction

The new English curriculum for Secondary Schools in Indonesia is a competency-based curriculum (CBC) requiring students to acquire ability in accomplishing tasks adequately, to find solutions, and to realize them in real situations. The curriculum instructs the school to assess the student's English competencies continuously. The English competencies that must be acquired are abilities of the students in using English as a means of social and interpersonal communication (Pedoman Khusus Model 3 Bahasa Inggris, 2002). Thus, the assessment of the students' English competencies definitely deals with the four English skills.

Such objectives can be achieved if the quality of education system as a whole, believing that competency is trainable, is guaranteed. Teachers, as one of the crucial factors of education system, must be qualified as they play a crucial role in achieving the school objectives. This means that the graduates of teacher training colleges (TTCs) must be competent in the field having thorough knowledge of English and skills in English; they should be communicatively competent not only in formulating grammatically correct sentences, but also in using sentences

to speak, listen, write, and read. English proficiency is required for they would be models for their students. Being teachers of English, they must be able to impart their competence to learners, meaning that they should possess language competence to a greater degree than that expected of their students.

Besides conducting good teaching-learning activities supported by good learning facilities and resources, to ensure graduating students' English proficiency, a test should be administered. Unfortunately, a standardized test which measures their English proficiency is not nationally but institutionally administered. Such a test varies in ensuring the English proficiency of the graduating students because the institutional English proficiency test is administered without any supervision from an independent body. As a result, graduates of TTCs often vary in their English competency; even many of them are not proficient enough to become teachers of English (Sadtono, 1995).

Considering such a condition, a proficiency English test focusing on the general knowledge or skills prerequisite to teach English in secondary schools, or exit from TTCs should be administered. Such a test should be administered independently so that a national standard of proficiency for teachers can be measured. However, administering a national English proficiency test is expensive and cumbersome as well. A solution to the problem is the application of computer adaptive tests (CAT). Given the rationale, application, and advantages of individual, standardized proficiency test using CAT on students with learning English difficulties in regular schools should be well considered as a positive development in language assessment practice.

Students with learning English difficulties in regular schools in this paper refer to specific category including students at risk (Friend and Bursuck, 2002: 27) and students with low motivation. Indonesian students learning English at TTCs have limited experiences in using English as a means of communication. Many of them even are not exposed to English outside the class activities so that they are more likely than others, who are exposed to English both in and outside the classrooms, to fail in school. In other words, they are students with limited English learning experiences. Including in the students at risk is "slow learners," whose English progress is slightly below average but who do not have a learning disability or who are not mentally retarded. Students with low motivation are the ones who may not have intrinsic motivation in learning English. They usually have to join the English department because of their parents' wish. Besides, they are usually not sure with the benefits of learning English.

The reasons for administering English proficiency of the graduating students of TTCs with learning English difficulties

It is widely believed that school achievement will improve if educations systems identify what is to be learned, then assess the

students' mastery of that material to determine the effectiveness of instruction. As such, an English curriculum administered to TTCs in Indonesia is almost the same: consisting both core and "local contents." The core curriculum consisting of English subjects is similar while the "local contents" are subjects which the local educational or institutional authority is free to develop; these local subjects are considered significant to equip students with skills necessary for their future profession. With such a curriculum, the graduates of TTCs are to be guaranteed. This argument may not be supported because other essential factors affecting the quality of the graduates are not considered, such as student's heterogeneity and assessment system.

Although an entrance test is administered to the candidates enrolling at TTCs, students of TTCs vary in their English proficiency. This condition still prevails, especially in the English classes of private TTCs, because the varied institutional entrance tests¹ do not calibrate student candidates' standard English proficiency. As a result, students' English proficiency varies among TTCs.

Input heterogeneity contributes to varied proficiency of English among students of TTCs. An assumption that the graduates of TTCs are proficient in English can be rejected for not only the heterogeneity among the students but especially due to the absence of reliable national standard calibration. Sadtono (1995) suspects that the pass or failure in the proficiency courses offered at TTCs depends on the individual lecturers and the norms and thus vary greatly. Even, for special cases, some students with limited English pass the course because the lecturer may be fed up with them and not, on the contrary to, let them have time and opportunity to improve themselves. In short, there is no quality control of the graduates of TTCs using objective external calibration.

An assumption that many teachers of English at high schools are not proficient enough to teach English (cf. Sadtono, 1995) is going to be stronger than before. Many secondary schools, where the graduates of TTCs teach, voice that many graduates of TTCs are not proficient to teach English. Although they graduated from TTCs, they are not proficient enough to become teachers of English. They have not yet attained levels of competence of English language proficiency that would, ordinarily, be deemed necessary and sufficient to allow them to practice as teachers of English. Even worse, many of them lack English knowledge and are less competent in the four language skills. While proctoring the graduating students of TTCs conducting a teaching practice at secondary schools in Surabaya, we often found teachers of English and the graduating students of TTCs hardly produce or understand a discourse of fluent and correct English. This condition is understood for up to the time being there is no

¹ Limited number of TTCs in Indonesia administer an English entrance test; most of TTCs administer an aptitude test instead of an English entrance test.

standardization of teacher trainees in Indonesia. Among teachers of English, issues on English testing are less important compared to those on curriculum and teaching techniques. Only in this year, 2003, “Developing Standards for the Teaching of English in response to the Global Challenges” becomes the main topic of the coming TEFLIN international conference.²

As teachers of English as a foreign language play a crucial role, as a model for language usage (correctness) and language use (meaningfulness), they are expected to be proficient in the language. To give correct and meaningful models to students, they should have a thorough knowledge of English; they have to understand how English works. Besides, they should be skillful in communicating in English both written and spoken. In line with this, Thomas (1987: 34) recommends that the main objective of teaching a language is to develop the learner’s mastery of it, it follows that the language teacher needs the competence to achieve it. Further, Sadtono (1995) concludes that language teacher competence is the competence to impart competence in language. An English teacher having little or insufficient competence in the language is not able to impart his competence to learners. In order to impart his competence to learners, an English teacher himself/herself should be competent to a greater degree than that expected of his/her learners.

English proficiency test for the graduating students of TTCs with learning English difficulties

Besides having poor language teaching and learning activities experienced in secondary schools, the graduating students of TTCs with learning English difficulties are often related to the students themselves, they are students at risk (cf. Friend and Bursuck, 2002). They usually have experiences that make them more likely than the others to fail in school. Students who live in poverty and lack learning resources or facilities or students with other social problems cannot often keep pace with the instruction in most regular classes. Some other students may tend to be “slow learners,” whose educational progress may be average but often fall behind the other normal motivated students.

Accepting students with learning English difficulties is actually a good policy seen from inclusive education. Including students with learning English difficulties in regular TTC will undoubtedly improve their opportunities to learn English courses and expand interests as students without learning English difficulties. This program not only benefits students with learning English difficulties but also provides a variety of benefits for students without learning English difficulties in

² TEFLIN stands from The Association of the Teachers of English as a Foreign Language in Indonesia. It is a non-governmental organization; its major activity is organizing workshops and seminars in English language teaching.

gaining valuable English skills and experiences required to face the challenges of a diverse and changing society. The issue will accordingly deal with how students succeed in their study in the regular schools and not with the nature of the quality of the student inputs.

In relation to assessing students' English proficiency, a question of opportunities for students with learning English difficulties will be crucial. Answers to this question will result in the implications of including students with learning English difficulties in statewide assessments. Further, there will be changes in organizational support, professional development, and educational practice that will make it possible for all students to succeed. Traditional paper-pencil testing, for example, will be challenged to meet the need of providing integrated information sources for students with learning English difficulties. In line with this, Friend and Bursuck (2002: 275) explain that assessment for students with learning English difficulties (special needs) requires a number of information sources including group-administered and individual-administered tests. Group-administered tests are used to assess students' English competence in general while individual-administered tests can be used to diagnose the students' English difficulties.

Another consequence of including students with learning English difficulties is consideration of test accommodation, i.e., alterations in the way a test is administered or the way a student provides responses that are designed to redress the student's disability. According to Warger (2002) appropriate accommodations do not provide an unfair advantage. Traditional paper-pencil testing is usually administered in a formal condition regardless of accommodation for students with learning English difficulties. All students taking paper-pencil tests are given the same opportunity. In paper-pencil tests, there is no adaptation in test construction for students with learning English difficulties. Meanwhile an adaptation in test construction, i.e. test items that are written clearly and assess pertinent knowledge or skills, is needed as students with learning English difficulties have benefited from it (Friend and Bursuck, 2002: 379). Another test accommodation is setting inclusion of which it is not well considered in traditional paper-pencil testing. Paper-pencil tests are usually administered in the regular classrooms where they might not give a pleasant learning experience. In other words, traditional paper-pencil testing may not accommodate the needs of students with learning English difficulties.

Referring back to the main issue of this paper, we conclude that English proficiency test should consider varied information, test modifications, and test accommodations appropriate for the graduating students of TTC with learning English difficulties. Tests administered to students with learning English difficulties, for example, are expected to provide information about general English competence as well as English difficulties. Types of test modifications focus, for example, on how responses are given with a minimally demanding language task (multiple-

choice format) versus a maximally demanding language task (short-answer format). Types of accommodation include changes in the time allocated for the test, when or where the test is given, presentation of the test (how the assignments are given), and how the student responds.

CAT as a solution?

Three crucial issues in inclusive tests are the inclusion of varied information, modification and accommodation policies for students with learning English difficulties. CAT can answer all these issues. Different from a conventional paper and pencil test which only provides the students' achievement ranks, CAT measures the student's achievement rank in the group as well as reflects his learning progress. This is possible for CAT is a computer-administered exam that adjusts to the ability level of the student taking the test as each question is presented (Brown, 1977; Dunkel, 1999). The advantage of this type of examination is that it is more targeted to the student's skill level and therefore can provide a more accurate estimate of the student's ability and can accomplish that goal using fewer questions.

Besides providing information about the ability levels of individual students, CAT enables the testers to have information on the student's English difficulties. This means that CAT could be used as a proficiency test and a diagnostic test as well. Serving as a diagnostic test, CAT can even provide diagnostic feedback very quickly to each student not only in the form of a report test scores but also on the items answered incorrectly (Brown, 1997:6). Further, CAT can provide information on the students' status: students who have no English learning difficulties and those who have learning English difficulties. This classification, however, cannot be used to compare the students' English achievement as CAT deals with individual student's English achievement.

For the questions to adapt to the level, CAT includes a large bank of test questions³, covering all levels of ability from basic to advanced. The questions are displayed on-screen, and the students respond using a keyboard and a mouse. As each question is answered, the computer assesses the response and selects the next question. The CAT starts by presenting the student with a question of "average difficulty". The next question is selected based on the student's response to the first question. If the student answers the first question correctly, the CAT will present the student with a slightly more difficult question. Conversely, if the

³ CAT has been made possible by the extensive use of Item Response Theory, and the development of algorithms that drive the test program to select and deliver test items, score responses, and provide immediate feedback to examinees. This paper does not discuss IRT methods for they are beyond the scope of this paper.

student answers a question incorrectly, the CAT will present the student with a slightly easier question. The remaining questions are selected in much the same way. Thus, the questions are selected according to whether the students' previous answer was right or wrong; in this way, the questions become progressively easier or harder until the system has a reliable assessment of the student's level.

Besides providing information about the levels of the student's proficiency, CAT can identify the levels of support in English which he needs to carry out his study effectively. Such varied information may not be provided by paper-and-pencil tests which are typically "fixed-item" tests in which the students answer the same questions within a given test. In the paper-and-pencil tests, all students are measured with items that are either very easy or very difficult for them. Accordingly, paper-and-pencil tests provide relatively little information about the students' ability level.

With the right large item bank and a high student ability variance, CAT can be much more efficient in administration, scoring, data analysis and data management, and score reporting than a traditional paper-and-pencil test (Rudner, 1998). Of the same opinion, Douglas (2000) argues that CAT is efficient since it presents test items which are at or near each test taker's level of ability from a bank of items of known difficulty. He further states that CAT offers advantages in terms of accessibility, scoring efficiency, and overall testing time.

Besides providing varied information, CAT may offer students with appropriate test accommodation. According to Stansfield et al (2001), accommodation for students with limited English proficiency include offering extra time and providing comfortable environment that allow the examinees to feel more comfortable, such as allowing them to take the test in a familiar setting and/or permitting a flexible schedule that includes shorter test sessions or more breaks. Similarly, Thomson et al. (2003) generate four categories of common accommodation: presentation, response, time/scheduling, and setting. These four are considered by most states in the U.S. when a test is administered to students with limited English proficiency. CAT, unlike conventional paper-and-pencil testing, is flexible, in terms of individualized timing; it can be administered at various time when students feel convenient. In addition, it can be administered outside the classroom as far as there is a computer (terminal or personal computer) connected to the Internet, for example, in the SAC (self-access center)⁴. To meet presentation accommodation, on the screen of a computer, students can enlarge the font of the letters and even color

⁴ SAC is a place where students can learn and assess themselves outside the class without the teachers. They can choose among different kinds of learning materials and with the help of the SAC counselors, they can interact to enhance different aspects of language proficiency (grammar, reading, listening, writing, speaking, and vocabulary). More and more English Departments in Indonesia consider and establish SAC.

the text to make them more readable and interesting. At last, CAT provides options for marking responses—mouse click, keyboard, touch screen, speech recognition—to the test items that are not available on paper (Thomson et al., 2003).

Some other advantages of CAT as a standardized proficiency test

In addition to the issues related to inclusive tests, advantages offered by CAT can broadly be classified into two classifications: testing considerations and human considerations (Brown, 1997). In terms of testing considerations, the advantages offered by CAT are accuracy at scoring selected-response tests and reporting score results, immediate feedback in the form of a report of test scores, minimum practice effects, and provision for diagnostic feedback. Compared to paper-and-pencil tests, CAT is more effective in terms of human consideration, i.e. allowing students to work at their own pace, taking less time to finish, experiencing less frustration in working out the test items (the students work on test items appropriate for their own ability levels), experiencing less overwhelming of test item presentation. These advantages are in line with the major features of a standardized test: uniform administration and some form of calibration (cf. Brown, 2001).

The distinguishing features of a standard test are uniform administration and some form of calibration. With advanced computer technology and extensive use of IRT, CAT meets these features. Test items included in the CAT can be both receptive-response items like multiple-choice, true-false and matching items, and productive-response item types such as fill-in and cloze (Brown, 1977; Dunkel, 1999). Such test forms are fairly easy to adapt to the computer-assisted testing medium, of which CAT is one of the type, and easily scored only as right or wrong.

The main issue in large-scale language proficiency testing is still in search of test methods to accurately assess language learners within the limits of practicality (Brown, 2001). In order to test multitudes in one administration, instruments need to be designed to mirror language tasks of the real world yet allow for rapid scoring at a marketable cost.

Today, with commercial CAT programs, such a test administration can be coped with. Efficiency could be gained for the developer only creates English CAT using software templates rather than developing CAT from the beginning. With objective test formats assessing language elements and skills as developed by Dunkel (1997) and Educational Testing Service and TOEFL (1999), CAT can answer issues in large-scale language proficiency testing.

The Crucial Issues of CAT

Although CAT shows promise in becoming a regular component of standardized foreign language assessment in the near future, a number

of questions need to be addressed when considering the basic principles of assessment in the CAT. Dunkel (1999) identifies the crucial issues of CAT as clear guidelines, specifications, and blueprints of CAT. These issues need to be considered not only by the developers but also the users of commercially developed CAT. Developing CAT is of course more expensive than using available commercially developed CAT. Commercial programs are now available such as the ones offered by the Assessment Systems Corporation (St. Paul, Minnesota) and Computer-Adaptive Technologies (Chicago, Illinois). These programs, according to Dunkel (1999), make it easier for developers to create English CAT using software templates rather than having to start programming and development from scratch.

Besides guidelines, specifications, and blueprints of CAT, Dunkel (1999) emphasizes issues involving the basic principles of assessment in the CAT. Special attention is to be given to general principles or criteria of a good test as the purpose of measuring, test reliability, and test validity (cf. Heaton, 1990; Weir, 1993; Brown, 2001). Considering that CAT can be used for a wide variety of purposes, English CAT developers need to clearly specify the assessment purpose of their tests as the specific objectives of English courses, the level of achievement in a language skill and structure skill domain, and the student's skill-area strengths and weaknesses. To ensure that CAT is able to measure the student's true English proficiency level, an English CAT must provide students with a sufficiently broad range of English content and skill tasks. In line with the topic of this paper, the developed English CAT should concentrate on assessing the students' general command of English for teaching or helping high school students learning English as a foreign language covering both language elements and language skills.

A test is reliable if it is consistent and dependable. A reliable test should be a consistent measure of performance. This means that the scores obtained by a student on one occasion should be very similar to those which would have been attained by the same student with the same ability if the test had been administered on a different occasion, if it were possible to bring such a situation about (which it isn't). The general factors influencing reliability that Dunkel (1999) identifies include, for example, whether test instructions are clear and explicit, or whether the students are amply familiar with the format of CAT before taking it. Besides, the situational factors, such as noise level, and individual factors as the students' psychological health and students' experience with similar tests influence the reliability of CAT.

By far the most complex criterion of a good test is validity, the degree whether a test actually measures what it purports to measure. There is no final, absolute and objective measure of validity. However, the validity of the test can be established by observation and theoretical justification (Brown, 2001). The test developer should be sure that the test items are the ones that reflect realistic use of the particular ability to be

measured, in this case the ability refers to the general command of English of a teacher of English as a foreign language in Indonesia. This content validity is not the only validity aspects that CAT developer should consider. According to Dunkel (1999), the aspects of validity that should be seriously taken, besides content, are construct, criterion, concurrent, and predictive.

The last substantial information that Dunkel (1999) thinks seriously is an introduction to the computer, the CAT system, the structure, organization, and content domains of the CAT. Considering that CAT is relatively a new testing system, the graduating students of TTCs should be given the time and opportunity to become thoroughly familiar with both the computer and the testing system. If they are not familiar with computer, they are given time for orientation of the functioning of the computer. The activities could be, for example, using a mouse, calling for questions, answering questions, adjusting the audio volume, and scrolling.

Conclusion

This paper discusses the promise of CAT as a standardized English assessment, particularly for assessing the graduating students of TTCs with learning English difficulties in Indonesia. Many benefits accrue to the graduating students of TTCs and administrators alike when using CAT. The benefits are not only limited to the general benefits that have been identified by scholars, such as Brown (1977), Dunkel (1999), but also provisions of accommodation for students with learning English difficulties. Such benefits could improve the proficiency of the graduating English students of TTCs, of which their English proficiency vary. To ensure that CAT has qualities of a good test, crucial issues related to CAT development must be well-considered. As Indonesia is a big country having many TTCs spreading in many cities with different condition and qualities, we think it is a time for the government to consider the development and use of CAT as an alternative testing system, besides the traditional paper-pencil testing system.

References

- Brown, J.D. 1977. Computers in Language Testing: Present Research and Some Future Directions. *Language Learning & Technology*. Vol. 1, No. 1, July 1997, pp. 44-59. <http://polyglot.cal.msu.edu/llt/vol1num1/brown/default.html>
- Brown, H. Douglas. 2001. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. London: Longman.
- Douglas, D. 2000. *Assessing language for Specific Purposes*. Cambridge: CUP.

- Dunkel, Patricia A. 1997. Computer-adaptive testing of listening comprehension: A Blueprint for CAT development. *The Language Teacher Online*, 21 1-8.
<http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/97/oct/dunkel.html>.
- Dunkel, Patricia A. 1999. *Considerations in Developing or Using Second/Foreign Language Proficiency Computer Adaptive Tests*. (2) 2, January, pp. 77-93.
<http://ilt.msu.edu/vol2num2/dunkel/default.html>
- Friend, M. and Bursuck, W.D. 2002. *Including Students with Special Needs: A Practical Guide for Classroom Teachers*. Boston: Allyn & Bacon.
- Fulcher, Glenn. 2000. Computers in Language Testing. In Brett, P. and G. Mottersm (eds) 2000. *A Special Interest in Computers*. Manchester: IATEFL Publications, 93 - 107.
<http://www.dundee.ac.uk/languagestudies/ltest/ltrfile/Computers.html>
- Heaton, J.B. 1990. *Classroom Testing*. London: Longman.
- Rudner, Lawrence M. 1998. *An On-line, Interactive, Computer Adaptive Testing Tutorial*. ERIC Digest.
<http://ericae.net/scripts/cat/catdemo.htm>
- Sadtono, E. 1995. *The Standardization of Teacher Trainees in Indonesia*. A Paper read at the Second International Conference on Language in Development: The Stakeholders' Perspective, Denpasar, Bali.
- Stansfield, C., and Rivera, C. 2001. *Test Accommodations for LEP Students*. ERIC Digest.
http://www.ericfacility.net/databases/ERIC_Digests/ed458289.html
- Thomas, A. L. 1987. *Language Teacher Competence and Language Teacher Education* .in ELT Document 125.
- Thompson, S., Thurlow, M & Moore, M. 2003. Using Computer-based Tests with Students with Disabilities. In *The National Center on Educational Outcomes*. No. 15 / January 2003.
<http://education.umn.edu/nceo/OnlinePubs/Policy15.htm>
- Warger, Cynthia. 2002. *Helping Students with Disabilities Succeed in State and District Writing Assessments*.
http://www.ericfacility.net/databases/ERIC_Digests/ed463622.html
- Weir, C. 1993. *Understanding & Developing language Tests*. New York: Prentice Hall.