An Investigation of Content Validity in O-NET (English Subject) for the Upper Secondary Level (Matthayom 6)

Nelobol Nipakornkitti and Dumrong Adunyarittigun

Faculty of Liberal Arts, Thammasat University

Abstract

The O-NET (Ordinary National Educational Test) for the upper secondary level (Matthayom 6) plays the role of accountability as well as gatekeeping in the Thai educational system. It has been used to hold schools accountable, and to be one of the criteria to determine students who are qualified for admission to universities. The results of the test have a great impact on stakeholders and especially on Thai students. Therefore, the O-NET test must be rigorously aligned with the Basic Education Core Curriculum in order to ensure a high degree of test validity. However, little is known about the validity of the test. Thus, the purpose of this study was to investigate content validity of the O-NET tests in the subject of English for the upper secondary level by using the test-curriculum alignment method. The test samples used in this study were the 2009 and 2010 O-NET tests. Five participants were purposively selected to match the O-NET test items with the Basic Education Core Curriculum. The results demonstrated that the O-NET test items were partially aligned with the national curriculum. The alignment between the test items and the national curriculum as well as the distribution of the test items on the curriculum domains are discussed.

Keywords: O-NET test, standardized test, high-stakes test, content validity, test-curriculum alignment

การตรวจสอบความตรงเชิงเนื้อหาของข้อสอบโอเนต วิชาภาษาอังกฤษชั้นมัธยมศึกษาปีที่ 6

นีโลบล นิภากรกิตติ และดำรงค์ อดุลยฤทธิกุล
คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์

บทคัดย่อ

ข้อสอบโอเนต (O-NET) (การทดสอบทางการศึกษาในระดับชาติขั้นพื้นฐาน) สำหรับมัธยมศึกษาชั้นปีที่ 6 มีบทบาทสำคัญในการสะท้อนความรับผิดชอบของผู้มีส่วนได้ส่วนเสียในระบบการศึกษาและเป็นกุญแจสำคัญที่นำไปสู่โอกาสทางการศึกษาของนักเรียนชั้นมัธยมศึกษาปีที่ 6 กล่าวคือ ผลสอบโอเนตถูกใช้เพื่อเป็นตัวชี้วัดประสิทธิภาพในการทำหน้าที่ของโรงเรียนและใช้เพื่อเป็นตัวกำหนดผู้มีสิทธิ์เข้าศึกษาต่อในระดับอุดมศึกษา ผลการสอบโอเนตจึงมีผลกระทบอย่างมากต่อผู้เกี่ยวข้องกับการสอบ โดยเฉพาะอย่างยิ่ง ต่ออนาคตของนักเรียนไทย ดังนั้น ข้อสอบโอเนตจึงต้องมีความสอดคล้องกับหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน เพื่อให้เกิดความมั่นใจว่าข้อสอบโอเนตมีความตรงสูง อย่างไรก็ตาม ข้อมูลเกี่ยวกับความตรงของข้อสอบโอเนตนั้นมีน้อยมาก ดังนั้น งานวิจัยชิ้นนี้จัดทำขึ้นเพื่อสะท้อนให้เห็นถึงความตรงเชิงเนื้อหาของข้อสอบโอเนตวิชาภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 6 โดยใช้วิธีการหาความสอดคล้องระหว่างข้อสอบกับหลักสูตร ตัวอย่างข้อสอบที่ใช้ในงานวิจัยนี้เป็นข้อสอบโอเนตปี 2553 และ ปี 2554 อาสาสมัครจำนวน 5 คน ได้รับการคัดเลือกอย่างเจาะจงเพื่อที่จะจับคู่ข้อสอบโอเนตกับหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ผลจากการวิจัยแสดงให้เห็นว่าข้อสอบโอเนตมีความสอดคล้องกับหลักสูตรแกนกลางเพียงบางส่วน อีกทั้งผลจากการวิจัยแสดงผลการกระจายตัวของคำถามในข้อสอบที่สอดคล้องกับประเด็นต่าง ๆ ในหลักสูตรแกนกลาง

คำสำคัญ ข้อสอบโอเนต แบบทดสอบมาตรฐาน ข้อสอบที่มีเดิมพันสูง ความตรงเชิงเนื้อหา ความสอดคล้องกันระหว่างข้อสอบกับหลักสูตร

1. Introduction

O-NET (Ordinary National Educational Test), known as a Thai state-mandated test, is designed based on the framework of the 2008 Basic Education Core Curriculum which defines the learning standards and indicators for all academic compulsory levels. The O-NET test for the English subject for the upper secondary level (Matthayom 6) was developed based on four domains: Culture, Communication, Connection, and Communities, known as 4Cs (Foley, 2005) which are prescribed in the 2008 Basic Education Core Curriculum with the objective to improve Thai students' proficiency in a competitive, globalized world.

The English O-NET (upper secondary level) test has been designed based on the 2008 Basic Education Core Curriculum which covers four important curriculum domains: language for communication, language and culture, language and relationship with other learning areas, and language and relationship with communities and the world. It consists of three main parts which are language use and usage, writing ability, and reading.

The O-NET test for the upper secondary level plays two important roles in the Thai educational system: holding accountability (Smith and Fey, 2000) and gatekeeping for students at the tertiary level (Wall, 2000). The O-NET for the upper secondary level has become an important tool to hold stakeholders accountable with the main purpose to improve the quality of teaching and learning in the Thai educational system. It is used to hold the stakeholders accountable: students, teachers, school administrators, and policy makers. Teachers and school administrators are expected to provide a high quality of education and motivating learning environment for all students. They are accountable to students who get service from them and parents who pay for service. Then, the achievement of students on the O-NET test should reflect the quality of education the students receive. It reveals how well teachers teach their students and help them develop their learning performance to meet the curriculum requirements and how well school administrators manage their school, teachers, and resources to make schooling successful. Besides, the O-NET for the upper secondary level also functions as a gatekeeper for students at the tertiary level. The test result is used as a criterion to recruit students into colleges (Wall, 2000).

Interestingly, the English O-NET (upper secondary level) test scores of Thai students in our educational system shockingly keep falling below standard. The test scores for the upper secondary level were 25.35 in 2012, 23.44 in 2013 and 24.98 in 2014 (www.niets.or.th). An important observation about the decline of the test results was whether it is due to the test itself or the accountability of teachers and school administrators. An interesting study by Lincharearn and her colleagues (2009) showed that one of the most important factors causing the poor O-NET test scores is the O-NET test itself. The level of difficulty in the O-NET test (English subject) for the upper secondary level is much higher than that taught in the classroom. However, it should

be noted that we cannot make a claim that the O-NET test has become less valid from the result from the previous study (Lincharearn et al., 2009). That result cannot be used as empirical evidence to prove that the O-NET test content for the upper secondary level (English subject) was incongruent with the national curriculum. The main reason to account for this is that the researchers implemented only the in-depth interview as a single method to collect data from the stakeholders which was relatively subjective and did not further examine the content validity of the test. An examination of the congruence between the O-NET test items and the national curriculum is needed in order to prove that the English O-NET test for the upper secondary level is rigorously constructed with validity (Messick, 1989; Brown and Abeywickrama, 2010).

Therefore, the objective of this paper is to investigate the alignment between the O-NET test items for the upper secondary level (English subject) and the national curriculum in order to fill the gap of content validity for the English O-NET test for the upper secondary level.

2. Theoretical Framework

Validity is considered as the most important element of a good test in order to confirm that the test results are meaningful enough to make major educational decisions. A valid test can provide appropriate, meaningful, and useful information for assessment (Gronlund, 1998). Content validity is a primary good test property which accounts for the boundaries of test content measured by test representatives and test coverage (Messick, 1994).

To enhance the degree of content validity, it is necessary to investigate the threats to validity which are construct-underrepresentation and construct-irrelevant variance (Messick, 1993, 1994, 1996, 1998). Construct-underrepresentation occurs when the test representatives and sample of domains are too narrow or too few when compared with the domains identified in the curriculum. In contrast, construct-irrelevance variance occurs when the test representatives and sample of domains are too broad and beyond the domains stated in the curriculum (Messick, 1989).

Therefore, to construct a valid test, it is necessary to decide whether the sample of the test items really represents the intended content to be measured (Brown, 2005;

Hughes, 2012). When test writers create a test, test items need to be rigorously congruent with test specifications (Brown, 2005; Hughes, 2012). Clear test specifications will guide test developers in writing more valid items. The test items need to be matched with the test specifications (Brown, 2005).

The examination of content validity is generally conducted by using the test-curriculum alignment method. In order to find evidence supporting content validity, the expert would justify whether the test items are congruent with the test specification (Messick, 1993; Webb, 1997; Bhola et al., 2003; Lopez, 2013). Mostly, “matching task” or “Likert-type rating scales” (Sireci and Faulkner-Bond, 2014) are used to measure the alignment between the test items and content domains. Certainly, the degree of congruence between the test items and curriculum domains needs to be validated by subject matter experts (Case, Jorgensen and Zucker, 2004; Sireci and Faulkner-Bond, 2014).

Yet, there are a few concerns about matching tasks. First, training the subject matter experts is the most important step (Li and Sireci, 2004; Martone and Sireci, 2009). If the subject matter experts are well trained with adequate sample tests and adequate amount of time, the matching results will be consistent because the subject matter experts will have a clear understanding of the criteria used to do the matching. Second, Bhola and her colleagues (2003) addressed that the over generosity of subject matter experts during the matching process can affect the matching results (cited in Li and Sireci, 2004). Therefore, an adequate training session will help minimize inconsistent results from matching.

3. Research Methodology

This study was conducted to investigate the extent to which the O-NET English test content for Matthayom 6 (upper secondary level) is aligned with the 2008 Basic Education Core Curriculum by using the test item-curriculum alignment method. In this study, the test specification was unavailable; therefore, matching test items with the curriculum domains was an effective method to provide empirical evidence to verify content validity of the test.

The participants in this study were purposively sampled. They were a researcher and four graduate students. Two of the graduate students were secondary school teachers, and the others were instructors at college level. They were qualified to be test item matchers since they had at least 5 years of teaching and testing experience at the secondary school level or college level, had taken a course in language assessment and measurement or test and measurement, and were not involved in developing O-NET test items for the secondary level.

The test samples used in this study were the 2009 and the 2010 O-NET tests. There were three O-NET tests (2008, 2009, and 2010 editions) officially publicized on the NIETS (National Institute Educational Testing Service) website (www.niets.or.th), at the time when this research was conducted. However, there was a limitation in using all of the three O-NET tests. That is, the 2008 O-NET English test for the upper secondary level was designed based on the former 2001 national curriculum whereas the 2009 and 2010 O-NET English tests for the upper secondary level were developed based on the 2008 Basic Education Core Curriculum. In addition, these two O-NET English tests, the 2009 and 2010 editions, had the same format. Each O-NET test consisted of 3 parts: language use and usage (10 test items), writing ability (20 test items), and reading (40 test items). One thing that made them different was the number of scores per item. The 2009 O-NET English test for the upper secondary level had 70 items for 70 points whereas the 2010 O-NET English test for the upper secondary level had 70 items for 100 points.

There were two instruments used in this study. The first instrument was a questionnaire designed for selecting qualified participants. The second instrument was a matching sheet devised for investigating the extent to which the O-NET English test for Matthayom 6 (upper secondary level) was aligned with the 2008 Basic Education Core Curriculum in terms of test content.

3.1 Data Collection

There were two crucial stages for examining the congruence between the O-NET test items and the 2008 Basic Education Core Curriculum: a preparation process and a matching process.

Preparation process

The five qualified participants (including the researcher) were introduced to the general information about the O-NET test and were trained how to consistently match the O-NET test items to the 2008 Basic Education Core Curriculum which consists of four curriculum domains (4Cs). The participants or matchers needed to have a clear understanding of the key words in the standard and the indicator (Ariyaritwikol et al., 2014, p. 222).

Table 1 gives information about the description of Standard F 1.2.3, its keywords and what students are expected to learn. The information about the standard is given in English and Thai.

Table 1: The key words in an indicator

The standard	The key words	What the student learns
Standard 1.2.3. Speak and write to express needs and offer, accept and refuse to give help in simulated or real situations. มาตรฐาน ต 1.2.3 พูดและเขียนแสดงความต้องการ เสนอ ตอบรับและปฏิเสธการให้ความช่วยเหลือในสถานการณ์จำลองหรือสถานการณ์จริงอย่างเหมาะสม	Speak and write พูดและเขียน	- พูดและเขียนภาษาที่ใช้ในการแสดงความต้องการ เสนอและให้ความช่วยเหลือ ตอบรับและปฏิเสธการให้ความช่วยเหลือในสถานการณ์ต่างๆ เช่น Please…/…, please./ I'd like…/ I need…/ May/Can/Could…? / Would you please…? Yes,… / Please do./ Certainly./ Yes, of course./ Sure./ Need some help?/ What can I do to help?/ Would you like any help?/ If you like I could…/ What can I do to help?/ Would you like any help?/ Would you like me to help you?/ If you need anything, please…/ Is there anything I can do?/ I'll do it for you./ I'm afraid…/ I'm sorry, but…/ Sorry, but… etc.

Furthermore, the matchers were trained to understand the objective of each test item and how to sort out each test item to make it correspond to each standard. They were asked to notice the key words in order to match each test item with an appropriate standard.

After training, the matchers practiced matching exercises together and then they independently matched the sample materials so as to enhance their understanding of the criteria for matching before they conducted the intended matching process.

Matching process

After the matchers independently matched the 2009 and 2010 test items to the 2008 Basic Education Core Curriculum for a week, they had a meeting in order to share and discuss difficulties and problems concerning what occurred during the matching task. Then, the matchers went back to recheck and revise their matching.

3.2 Data Analysis

As can be seen in Table 2, the results from the matching procedure were analyzed in order to determine an internal correlation among the five matchers. The content validity of the O-NET test was validated by the extent to which the O-NET test items were aligned with the 2008 Basic Education Core Curriculum.

Table 2: Inter-rater reliability among the five matchers

Parts of the O-NET test	Measured standard	Inter-rater reliability	
		2009	2010
Reading part	Standard F 1.1	1.0	1.0
Language Use and Usage part	Standard F 1.2	0.88	0.98
Writing part	Standard F1.3	1.0	1.0
	Overall	0.97	0.99

The degree of internal correlation among the five matchers was high. The inter-rater reliability was at 0.97 and 0.99 for the overall matching in the 2009 and 2010 O-NET tests of English (upper secondary level), respectively. According to Table

2, in both reading and writing parts of the 2009 and 2010 O-NET tests of English, the degree of inter-rater reliability was at 1.0. However, only in the language use and usage part, the degree of internal correlation among the matchers was a little lower than the other parts; that is, the 2009 edition was at 0.88 and the 2010 edition was at 0.98. The cutoff score of 0.5 or above was employed as a threshold to justify that the item was adequately aligned with the objective (Polit et al., 2007; D'Agostino et al., 2008).

The results of the internal correlation among the matchers indicate that the matchers were properly trained and that the matching done by the five matchers was highly consistent.

4. Results and Discussion

The results obtained from the matching process were analyzed. The 2009 and 2010 O-NET tests partially matched to the 2008 Basic Education Core Curriculum. The 2009 and 2010 O-NET test items were aligned with the curriculum domain of language for communication (curriculum domain 1). The results can be divided into two sections: matched curriculum domain and unmatched curriculum domains.

4.1 Matched curriculum domain

There were two crucial factors affecting the results of the alignment between the O-NET test items and the curriculum domain as follows. First, a test item should measure one objective at a time in accordance with language assessment principles (Brown and Abeywickrama, 2010). This made the matchers match a test item with the most appropriate standard. Secondly, the key words of the indicators in curriculum domain 1 (curriculum domain of language for communication) requiring students to perform language communicative skills can be measured by indirect assessment. Therefore, the O-NET test items which were designed to indirectly measure speaking, writing, and reading skills were matched with the standards of curriculum domain 1 measuring the students' language communicative skills.

For instance, Standard F 1.1 was designed to measure the students' understanding when they listened to and read a variety of texts, and were asked to provide opinions about what they listened to and read. The O-NET test items intended to measure Standard F 1.1 indirectly evaluated students' comprehension of what they have read in

the reading part. The other example is Standard F 1.2 requiring students to communicate and express their opinions about matters in their daily life in English. Dialogues appearing in the test were used to measure this standard. Students were evaluated on their ability to appropriately ask questions and provide responses appropriate for different contexts and occasions.

In addition to the results for the test-curriculum alignment, the matching results showed the distribution of O-NET test items among the standards which are shown in Table 3.

Table 3: The distribution of O-NET test items of English for the upper secondary level for each part of the O-NET test on curriculum domain 1

O-NET's part	**Curriculum domain 1: Language for communication**		**O-NET test items**	
			2009	**2010**
Reading	Standard F 1.1	Understanding and capacity for interpreting what has been *heard and read* from various types of media, and ability to express opinions with proper reasoning	50%	50%
Language Use and Usage	Standard F 1.2	Endowment with *language communication skills* for exchange of data and information; efficient expression of feelings and opinions	25%	25%
Writing	Standard F 1.3	Ability to present data, information, concepts and views about various matters through *speaking and writing*	25%	25%

As can be seen in Table 3 which presents the distribution of the 2009 and 2010 O-NET test items (English subject for the upper secondary level) of each part of the O-NET test on the curriculum domain 1, the percentages of the 2009 and the 2010 O-NET test items matching with Standard F 1.1 were similar. 50% of the test items in the O-NET test of English for the upper secondary level were congruent with Standard F 1.1 which can be considered as an indicator to measure the reading skill. Moreover, the results from the matching process showed that 25% of all the test items in each O-NET test were aligned with Standard F 1.2 as an indicator measuring language

communication skills. Finally, 25% of all test items of each O-NET test edition were matched with Standard F 1.3 which was an indicator measuring the speaking and writing skills.

The distribution of the 2009 and 2010 O-NET test items was aligned with the curriculum domain of language communication (curriculum domain 1) which was due to the fact that the 2009 and 2010 O-NET tests were possibly designed to measure receptive and productive skills at the same proportion. That is, as can be seen from the distribution on the 2009 and 2010 O-NET tests, 50% of the test items were meant to measure receptive skills through the reading part, and 50% of the remaining test items were designed to measure productive skills through the language use and usage part, and the writing part.

Hence, half of the O-NET test items were matched with Standard F 1.1.4 which required students to perform many sub-skills of reading, for instance, identifying main ideas, making inferences, skimming and scanning, and guessing meaning from context clues. In addition, the remaining test items (50%) of each edition were equally distributed to measure productive skills: speaking (25%) and writing (25%). Therefore, 25% of the O-NET test items in the language use and usage part were matched with Standard F 1.2, whereas 25% of the remaining O-NET test items in the writing part were matched with Standard F 1.3.

4.2 Unmatched curriculum domains

The results derived from the matching process revealed that the 2009 and 2010 O-NET tests of English for the upper secondary level were partially aligned with the 2008 Basic Education Core Curriculum. Some curriculum domains were not matched with the O-NET test items such as language and culture (curriculum domain 2), language and relationship with other learning areas (curriculum domain 3), and language and relationship with community and the world (curriculum domain 4) (see Appendix A).

When the content of the standards was analyzed, it was found that the key words used in the indicators require students to perform integrated skills (Hughes, 2012) and higher order skills, for instance, *read aloud* (Standard F 1.1.2), *analyze/ discuss* (Standard F 2.2.2), *research/ search for* (Standard F 3.1.1), *make records* (Standard F 3.1.1), *disseminate/ convey* (Standard F 4.2.2) and so on. This clearly

showed that these standards were designed to develop students' abilities to use integrated skills to communicate and to perform higher order skills in a competitive, globalized world (Ministry of Education, 2008). These cannot simply be measured by the O-NET test, which is an indirect assessment.

An example is Standard F 2.1.3 which requires students to "participate in, give advice and organize language and cultural activities appropriately". To evaluate students' ability to perform the aforementioned task, teachers have to observe how students make use of English to communicate with others in cultural activities or how they cooperate with others and do problem-solving when they do the task.

Another example is that Standard F 4.2.2 requires students to "disseminate/ convey to the public data and news about the school, community and the local area/ the nation in foreign languages". To illustrate, teachers may ask students to write announcements or news about cultural activities (e.g. Christmas Day activity, market day, or international day) that will be held in school. Students need to search for and collect information in English and interview informants in order to write news or announcements about the activities or the special events. As can be seen from the examples, it is difficult to assess students' abilities designated by the standards by using paper-and-pencil tests. Therefore, it was found that none of the 2009 or 2010 O-NET test items was directly aligned with the standards in curriculum domain 2, curriculum domain 3, or curriculum domain 4. It is evident that these standards require students to perform integrative tasks by using multiple skills and apply higher order thinking to do the activities in order to evaluate both process and product in a meaningful way. Consequently, these curriculum domains need to be evaluated by using performance-based assessment.

However, the O-NET test, a multiple-choice test, has some limitations to measure these standards in curriculum domains 2, 3, and 4 (O'Malley and Pierce, 1996). First, multiple-choice tests demand the use of discrete skills, not integrative. Second, multiple-choice tests cannot measure a wide range of higher order thinking skills. Third, using multiple-choice tests limits authenticity; that is, students cannot engage in real-world activities. The types of texts used to generate multiple-choice test items are not as complex or authentic as doing science projects or cultural activities. Finally,

multiple-choice tests emphasize product rather than process (Adunyarittigun, 2001). Because of the limitations of the multiple-choice test items in the O-NET test, it is clear that the standards in curriculum domains 2, 3, and 4 cannot be measured.

Hence, it is necessary to measure these curriculum domains using formative assessment or performance-based assessment which encourages students to perform integrative skills and make use of higher order thinking skills, encourages them to perform authentic tasks, and focuses on both product and process. In the guidelines for the measurement and evaluation based on the 2008 Basic Education Core Curriculum, the Office of the Basic Education Commission or OBEC (2011) promotes applying both formative and summative assessments at all levels of education. Therefore, it is suggested that more than half of curriculum domains requiring integrated language skills (Adair-Hauck et al., 2006; Hughes, 2012) be evaluated through performance-based assessment (O'Malley and Pierce, 1996).

5. Conclusion

5.1. Based on the results of the study, it was found that the 2009 and 2010 O-NET tests were partly aligned with the 2008 Basic Education Core Curriculum. The 2009 and 2010 O-NET tests were congruent with the curriculum domain of language for communication (curriculum domain 1).

5.2. This study revealed the content validity of the 2009 and 2010 O-NET tests. With the test-curriculum alignment method, the 2009 and 2010 O-NET test items were matched to the curriculum domain of language for communication (curriculum domain 1) due to the fact that the objectives of curriculum domain 1 were designed to measure students' communicative language skills: listening, speaking, reading and writing (Ministry of Education, 2008) which can be indirectly assessed by multiple-choice tests.

5.3. The results also showed that the 2009 and 2010 O-NET tests did not match to the other curriculum domains (curriculum domains 2, 3, and 4) because the objectives of the remaining curriculum domains require students to be able to communicate in English effectively through integration of skills and apply higher order thinking skills. Therefore, the remaining curriculum domains can be assessed by performance-based assessments in the classroom.

6. Implications and suggestions

The following implications and suggestions can be made from this study.

6.1. O-NET tests should be developed to cover four curriculum domains as much as possible so as to contribute to a higher degree of content validity. Moreover, test items should have adequate representatives and samples of the domains in order to make the tests have a higher degree of content validity. Because the O-NET test results are used to hold stakeholders accountable, it is necessary to develop a test theoretically based on language assessment principles (Brown and Abeywickrama, 2010), especially content validity which is considered as the primary characteristic of a test.

6.2. A test specification is needed to explore content validity of the O-NET tests. Test specifications can be used to verify the test writers' intention in assessing the students' abilities and to precisely indicate the degree of content validity of the O-NET tests. Yet, in case of not having test specifications, an interview of the test developers may help provide important information about the tests' objectives. Without specification of tests, it is difficult to identify the degree of content validity as discussed in this study.

6.3. The results of this study promote the application of both summative and formative assessments due to the fact that the national curriculum aims at developing students' abilities in many dimensions: four language communicative skills, integrative skills, and higher order thinking (OBEC, 2011). Teachers, parents, and stakeholders should be aware of the limitation of summative assessment and accept the use of formative assessment or authentic assessment in the classroom. This will help those involved to get accurate information about students' abilities, leading them to help students reach the goals of the national curriculum.

6.4. The procedures of collecting data and verifying the test used in this study can be applied to further study of content validity because they have been thoroughly verified and proved to be reliable via the matching process.

7. Limitations

The generalizability of the results of this study is limited because of the following:

7.1. This study was conducted based on two O-NET tests. The O-NET tests used in this study were 2009 and 2010 editions which were officially available on the NIETS website, and they were based on the framework of the 2008 Basic Education Core Curriculum. The latest editions of the test which are based on the 2008 Basic Education Core Curriculum have not been made available to the public. Thus, there is a need for further studies to investigate the alignment between the latest edition of the O-NET test items and curriculum.

7.2. In order to investigate content validity, there must be test specifications to cross-check the content that test writers aim to measure. In this study, the specifications of the 2009 and 2010 O-NET tests were not officially available to the public. Thus, the results of this study were analyzed based on the interpretation of the researcher without the test developers' test specifications.

เอกสารอ้างอิง/References

วรรณกวินท์ อริยฤทธิ์วิกุล, ชาติ แจ่มนุช, ไพเราะ มีบางยาง, นฤเทพ ใจสุทธิ และอิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2557). *การออกแบบการจัดการเรียนรู้ตามหลักสูตรอิงมาตรฐานหลักสูตรโรงเรียนมาตรฐานสากลและแนวการจัดการเรียนรู้สู่ประชาคมอาเซียน.* กรุงเทพมหานคร: สกสค. ลาดพร้าว.

ศึกษาธิการ, กระทรวง. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพมหานคร: กระทรวงศึกษาธิการ.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2554). *แนวปฏิบัติการวัดและการประเมินผลการเรียนรู้ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

เอื้อมพร หลินเจริญ, สิริศักดิ์ อาจวิชัย และภีรภา จันทร์อินทร์. (2552). *ปัจจัยเชิงสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ของนักเรียนชั้นประถมศึกษาปีที่ 6 และมัธยมศึกษาปีที่ 6 ต่ำ.* สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน).

Adair-Hauck, B., Glisan, E. W., Koda, K., Swender, E. B., & Sandrock, P. (2006). The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals, 39*(3), 359-382.

Adunyarittigun, D. (2001). Trends in EFL Reading Assessment. *Journal of Language and Linguistics, 19*(2), 69-75.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21-29.

Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment, Principles and Classroom Practices* (2nd ed.). NY: Pearson Education.

Brown, J. D. (2005). *Testing In Language Programs: A Comprehensive Guide To English Language Assessment.* NY: McGraw-Hill College.

Case, B. J., Jorgensen, M. A., & Zucker, S. (2004). Alignment in educational assessment. Retrieved June, 5, 2005, from www.pearsonassessments.com.

D'Agostino, J. V., Welsh, M. E., Cimetta, A. D., Falco, L. D., Smith, S., Van Winkle, W. H., & Powers, S. J. (2008). The rating and matching item-objective alignment methods. *Applied Measurement in Education, 21*(1), 1-21.

Foley, J. A. (2005). English in…Thailand. *Regional Language Centre Journal, 36*(2). 223-234.

Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Boston: Allyn & Bacon.

Hughes, A. (2012). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Li, S., & Sireci, S. G. (2004). *Evaluating the fit between test content, instruction, and curriculum frameworks: A review of methods for evaluating test alignment* (Center for Educational Assessment Research Report No. 558). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Lopez, Alexis A. (2013). Alignment between standardized assessments and academic standards: The case of the Saber Mathematics Test in Colombia. *RELIEVE, 19*(2). DOI: 10.7203/relieve.19.1.3026

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 79*(4), 1332-1361.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). NY: Macmillan.

_____. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series, 1993*(2).

_____. (1994). Alternative modes of assessment, uniform standards of validity. *ETS Research Report Series, 1994*(2).

_____. (1996). Validity and washback in language testing. *ETS Research Report Series, 1996*(1), i-18.

_____. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*(1), 35-44.

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Boston, MA: Addison-Wesley Publishing Company.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*(1), 100-107.

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334-344.

Wall, D. (2000). The impact of high-stakes testing on teaching on teaching and learning: can this be predicted or controlled? *System, 28,* 499-509.

Webb, N. L. (1997). Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. (*Research Monograph No. 6).* Council of Chief State School Officers: Washington, DC.

Appendix A: The 2008 Basic Education Curriculum (English subject for the upper secondary level)

Curriculum domain		**Indicators**
Curriculum domain 1: Language for Communication		
Standard F 1.1	Understanding and capacity for interpreting what has been heard and read from various types of media, and ability to express opinions with proper reasoning	1. Observe instructions in manuals for various types of work, clarifications, explanations and descriptions heard and read. 2. Accurately read aloud texts, news, advertisements, poems and skits by observing the principles of reading. 3. Explain and write sentences and texts related to various forms of non-text information, as well as specify and write various forms of non-text information related to sentences and texts heard or read. 4. Identify the main idea, analyze the essence, interpret and express opinions from listening to and reading feature articles and entertainment articles, as well as provide justifications and examples for illustration
Standard F 1.2	Endowment with language communication skills for exchange of data and information; efficient expression of feelings and opinions	1. Converse and write to exchange data about themselves and various matters around them, experiences, situations, news/incidents and issues of interest to society, and communicate the data continuously and appropriately. 2. Choose and use requests and give advices, clarifications and explanations fluently. 3. Speak and write to express needs and offer, accept and refuse to give help in simulated or real situations. 4. Speak and write appropriately to ask for and give data, describe, explain, compare and express opinions about matters/ issues/news and situations heard and read. 5. Speak and write to describe their own feelings and opinions about various matters, activities, experiences and news/ incidents with proper reasoning.

Curriculum domain		**Indicators**
Standard F 1.3	Ability to present data, information, concepts and views about various matters through speaking and writing	1. Speak and write to present data themselves/experiences, news/incidents, matters and various issues of interest to society. 2. Speak and write to summarize the main idea/theme identified from analysis of matters, activities, news, incidents and situations in accordance with their interests. 3. Speak and write to express opinions about activities, experiences and incidents in the local area, society and the world, as well as provide justifications and examples for illustration.
Curriculum domain 2: Language and Culture		
Standard F 2.1	Appreciation of the relationship between language and culture of native speakers and capacity for use of language appropriate to occasions and places	1. Choose the language, tone of voice, gestures and manners appropriate to various persons, occasions and places by observing social manners and culture of native speakers. 2. Explain/ discuss the lifestyles, thoughts, beliefs and origins of customs and traditions of native speakers. 3. Participate in, give advice and organize language and cultural activities appropriately
Standard F 2.2	Appreciation of similarities and differences between language and culture of native and Thai speakers, and capacity for accurate and appropriate use of language	1. Explain/ compare differences between the structures of sentences, texts, idioms, sayings, proverbs and poems in foreign languages and Thai language. 2. Analyze/ discuss similarities and differences between the lifestyles, beliefs and culture of native speakers and those of Thais, and apply them appropriately.
Curriculum domain 3: Language and Relationship with Other Learning Areas		
Standard F 3.1	Usage of foreign languages to link knowledge with other learning areas, as foundation for further development and to seek knowledge and widen one's world view	Research/ search for, make records, summarize and express opinions about the data related to other learning areas, and present them through speaking and writing.

Curriculum domain		**Indicators**
Curriculum domain 4: Language and Relationship with Community and the World		
Standard F 4.1	Ability to use foreign languages in various situations in school, community and society	Use language for communication in real situations/ simulated situations in the classroom, school, community and society.
Standard F 4.2	Usage of foreign languages as basic tools for further education, livelihood and exchange of learning with the world community	1. Use foreign languages in conducting research, collecting, analyzing and summarizing knowledge/ various data from the media and different learning sources for further education and livelihood. 2. Disseminate/ convey to the public data and news about the school, community and the local area/the nation in foreign languages.

(Ministry of Education, 2008)