

Assessment in Thai ELT: What do Teachers Do, Why, and How can Practices Be Improved?

Richard Watson Todd ^{a,*}, Duangjaichanok Pansa ^b, Natjiree Jaturapitakkul ^c, Nawiya Chanchula ^d, Punjaporn Pojanapunya ^e, Saowaluck Tepsuriwong ^f, Stuart G. Towns ^g, & Wannapa Trakulkasemsuk ^h

^a irictodd@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^b duangjaichanok.pan@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^c natjiree.jat@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^d nawiya.cha@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^e punjaporn.poj@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^f saowaluck.tep@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^g stuart.tow@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

^h wannapa.tra@kmutt.ac.th, Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

* Corresponding author, irictodd@kmutt.ac.th

APA Citation:

Watson Todd, R., Pansa, D., Jaturapitakkul, N., Chanchula, N., Pojanapunya, P., Tepsuriwong, S., Towns, S. G., Trakulkasemsuk, W. Assessment in Thai ELT: What do teachers do, why, and how can practices be improved? *LEARN Journal: Language Education and Acquisition Research Network*, 14(2), 627-649

Received
02/02/2021

Received in
revised form
09/04/2021

Accepted
20/05/2021

Keywords

assessment
practices,
positive
washback,
continuous
assessment,

Abstract

English language education in Thailand is rated poorly internationally and needs improvement. Research identifying the reasons for the I highlights assessment as a key concern since assessment has washback effects on learning. However, little is known about the assessment practices teachers use. This paper has three goals: to identify the assessment practices used, to identify the reasons underlying teachers' decisions about assessment, and to promote assessment practices likely to have positive washback. A survey of 329 teachers found that two-thirds of students' scores came from tests, but that all teachers used some form of continuous assessment. From interviews with 29 teachers, teachers used tests to reduce their workload and to prepare students for the national ONET exams, while using continuous assessment to motivate students and to prepare them for real-world uses of English. Based on

English language education, Thailand	evidence that continuous assessment is more likely to promote positive washback, 303 assessment activities designed by teachers were analyzed for potential benefits. Rating the activities on five criteria, 39 assessment activities with potential positive washback were identified, designed into attractive formats, and disseminated to teachers. Using these activities with positive washback is likely to have beneficial effects on English language education in Thailand.
--------------------------------------	--

1. Introduction

Thailand has gained an unfortunate reputation for the quality of its education, especially English language education. Evidence for this can be found in Thailand's low placement on international rankings of countries for education. Following Watson Todd (2015) in using normalized rankings where 0 is the lowest ranked country and 1 is the highest, some recently published reports give Thailand rankings of 0.20 for overall education quality, 0.22 for education results, 0.30 for language skills (all from IMD, 2020), 0.29 for performance on the TOEFL® test (ETS, 2020), and 0.26 for English language proficiency (English First, 2019). These rankings show Thailand as being in the bottom 30% of countries on several different aspects of education.

There have been numerous attempts to identify the reasons underlying Thailand's poor performance in English language education. For example, a report from Chulalongkorn University in 2000 (see Foley, 2005) identified several contributing factors including too much content in the curriculum, teachers' heavy workload, inadequate teaching media, large class sizes, and too many multiple-choice tests. Similar lists can be found in Thongsri et al. (2006) and Noom-Ura (2013). The last of these factors, assessment, is rated as a serious problem in these studies and is also shown to affect teachers' classroom behaviors (Fitzpatrick, 2011). The most recent UNESCO report on Thai education (OECD, 2016) also identifies an over-reliance on testing focusing on "the reproduction of discrete knowledge" (p. 165) as a key problem, and proposes a move towards a broader mix of assessments as a solution. This report, however, adds the caveat that there is no clear data available on the actual assessment practices of Thai teachers. One of the goals of the current study is to provide such data.

Given the likelihood that assessment practices are at least partly responsible for Thailand's poor educational performance and that little is known about teachers' actual assessment practices, our first goal in this paper is to identify what assessment practices teachers use in schools. Based on washback models of how student assessment affects teaching practice (see Cheng, 2014), from these results we may be able to posit the impacts of the assessment practices. Especially if teachers are using assessment practices with potentially negative effects, it would be useful to know why teachers choose to use the practices that they do. Therefore, our second goal is to report the reasons underlying teachers' decisions about assessment. Our third goal is to promote positive practices by collecting and disseminating model assessment activities which are likely to have positive washback. Given that each of these three goals requires different methodologies, in this paper we will report the research as three separate sequential studies.

2. Review of Literature

2.1 The Impacts of Assessment

Assessments can serve a wide variety of purposes. The Thai Basic Education Core Curriculum (Basic Education Commission, 2008) lists the main purposes as informing students of their progress, providing information for improvement, providing input into policies, and issuing documents of educational qualifications. The last of these serves Brown's (2013) gatekeeping function of assessment which appears to dominate in Thailand with assessment results underpinning decisions to allow students to progress through the various levels of education and gain employment. The prioritizing of the gatekeeping function has two major impacts. First, students start believing that gaining the highest marks on assessments is more important than actually learning (Watson Todd, 2008). Second, assessments with high reliability are valued over those with high learning validity (Tomlinson, 2005).

The high value placed on reliability of assessments leads to types of assessment for which reliability levels can be clearly demonstrated, such as multiple-choice tests, being preferred over types of assessment for which the reliability is unclear, such as developmental portfolios. The use of these different types of assessment can have different impacts on the teaching and learning process. These effects of assessment on teaching and learning are termed washback (see Brown, 2005).

Washback from assessment to teaching and learning can apply at all levels. The impact of national-level assessments in Thailand, such as the ONET exams (Ordinary National Education Tests), has been the focus of a substantial amount of research, most of which highlights negative impacts (Prapphal, 2008). The national-level ONET exams comprise multiple-choice items only and thus their impact can be monitored fairly clearly. The potential impacts of multiple-choice testing in ELT include focusing on language knowledge (i.e., vocabulary and grammar), an emphasis on receptive skills, promotion of rote learning, and a shift away from higher-order thinking (Brown, 2005; Burke, 1999). There is some evidence that the ONET exams have these negative washback effects (Nonthaisong, 2015). In addition, the national-level exams are often used as models for school assessments in Thailand (Watson Todd, 2008), and thus their impact may be even more far-reaching. The emphasis on receptive skills encouraged by the ONET exams means that writing and speaking are under-represented in school assessments creating negative washback (Jianrattanapong, 2011; Poonpon, 2012). Although much of the research into the impacts of assessments in Thailand provides a bleak picture, teacher-made assessments for use in their own classrooms can also have washback, and these assessments can have positive washback effects leading to beneficial teaching and learning practices. For example, several types of progressive continuous assessment (by which we mean non-test assessments which are likely to have positive washback impacts) have been shown to support more effective student learning (Cheng, 2014).

This brief summary of washback in Thai ELT is somewhat simplistic as not all multiple-choice testing is necessarily negative, not all continuous assessment is necessarily positive, and there are numerous other factors in addition to washback effects that need to be taken into account in making decisions about assessment (Spratt, 2005). Nevertheless, a context with an over-reliance on multiple-choice testing and under-use of progressive continuous assessment is

likely to create negative washback. It is therefore important that we know what assessments are being used in Thai schools to see what the potential washback effects are.

3. What Assessments Are Used in ELT in Thai Schools? (Study 1)

As noted in the UNESCO report on Thai education, little is known about how teachers actually assess their students' learning in Thai schools. There is, however, a plethora of advice on the issue. Both the National Education Act of 1999 and the Core Curriculum recommend that teachers should use a variety of methods with the former suggesting that students should be assessed "through observation of their development; personal conduct; learning behaviour; participation in activities and results of the tests" (Office of the National Education Commission, 1999, Section 26). Some research shows that teachers are aware of these recommendations but still predominantly use multiple-choice tests (e.g., Nonthaisong, 2015).

Previous research investigating English teachers' assessment practices at Thai schools in depth are Piboonkanarax (2007) and Narathakoon et al. (2020). Surveying the practices at 78 secondary schools, Piboonkanarax (2007) found that exams accounted for 60% of secondary school scores on average, with 90% of the marks on these exams coming from multiple-choice items. Progressive continuous assessment practices, although widely used, made minor contributions to students' overall scores. For example, portfolios accounted for 5% of scores on average. Focusing on primary school teachers, Narathakoon et al. (2020) asked teachers to rate how frequently they used various assessment methods and how effective they believed them to be. The most frequently used methods were final exams and mid-term exams, both of which were rated highly for effectiveness as well.

Since multiple-choice tests often have negative washback impacts, Piboonkanarax's findings are worrying. However, both the context and the teachers' practices may have changed in the time since the survey was conducted. The first goal of this article, therefore, is to replicate Piboonkanarax's study to see what ELT assessment practices are currently used in Thai schools.

3.1 Methodology (Study 1)

To identify teachers' assessment practices, a questionnaire similar to that of Piboonkanarax was designed. The questionnaire consists of three sections: background information (such as location and level of school), details of how scores are assigned for one course, and details of one assessment activity (this section was used to address the third goal of this study). The questionnaire is available at shorturl.at/hDTU6. The course score data is the main data source for the first research goal with the background information analyzed as variables potentially influencing how scores are assigned. Institutional ethics approval was obtained for conducting the survey (as well as for the other stages of the research project).

E-mails were sent to 4,020 English language teachers at both primary and secondary schools throughout Thailand, and 347 responded, of which 329 responses were usable (the discarded responses included inconsistent or invalid information). These return rates are reasonable for unsolicited e-mails from unknown senders. Frequencies and means were calculated for the different assessment types. Point biserial correlations were calculated to

compare the percentages assigned to the broad assessment types with some of the background variables using a probability value of less than 0.05 as significant.

3.2 Results (Study 1)

In focusing on the likely washback impacts of the assessment practices in Thai schools, the issues of particular relevance are:

1. The proportions of scores assigned to various broad assessment types (i.e., final exams, mid-term exams, quizzes, and continuous assessment).
2. The proportions of scores assigned to various broad assessment approaches (i.e., closed assessment where there is a single correct answer, open assessment which is usually productive, and non-linguistic behavior such as participation).
3. The proportions of scores assigned to different task formats (e.g., multiple-choice, gap-fill, role-plays, portfolios). Those task formats used by at least half of all teachers are reported.

For each of these we can report the proportion of teachers who use each category and the proportion of the total score for a course assigned to each category. These findings are shown in Table 1.

TABLE 1
Teachers' Use of Different Categories of Assessment

Category of assessment	% of teachers using this category	% of total score assigned to this category
<i>Broad assessment types</i>		
Continuous assessment	100	35.6
Final exam	100	26.6
Quizzes	91	20.0
Mid-term exam	92	17.8
<i>Broad assessment approaches</i>		
Closed assessment	100	58.1
Open assessment	94	27.1
Non-linguistic behavior	90	5.3
<i>Task formats</i>		
Multiple-choice	97	37.4
Presentations	85	10.1
Homework	86	9.3
Reports/projects	75	9.0
Role-plays	75	6.8
Portfolios	77	6.5
Gap-fill	74	6.4
Matching	63	3.8
Short-answer	62	3.7
External reading	61	3.7

From a washback perspective, Table 1 presents a mixed picture. Worryingly, about two-thirds of the assessment comes from tests in the form of exams or quizzes, over 60% of the assessment measuring language performance is closed, and multiple-choice is by far the most

common item type. These categories of assessment are those most likely to promote rote learning and the teaching of simplistic knowledge and least likely to promote the productive skills and higher-order thinking. However, there are some positive results as well. All teachers use some form of continuous assessment and the majority use progressive assessment practices such as projects, portfolios and external reading. In other words, practices that are more likely to promote deep thinking, reflection and broad-based language goals are frequently used by teachers. These findings show some improvements since Piboonkanarax conducted her survey in 2007. Both surveys show test-dominated assessment schemes with high use of multiple-choice and some minor use of continuous assessment practices, but our survey shows a slightly greater use of assessment practices likely to be associated with positive washback than was the case in 2007.

Teachers responding to the survey worked at all levels of school education and came from around the country. Comparing their places of work with the percentages given to some of the most highly used key categories, Table 2 shows whether there are any differences in assessment practices between different types of schools.

TABLE 2
Patterns of Assessment in Different Types of Schools

Assessment category	r_{pb} (point biserial correlation)	p (probability)	Interpretation
<i>Comparison between primary (N = 131) and secondary schools (N = 189)</i>			
% from final exam	0.13	0.02	Secondary schools rely more on final exams
% from mid-term exam	-0.23	<0.01	Primary schools rely more on mid-term exams
% from continuous assessment	0.10	0.07	Not significant
% from multiple-choice	-0.19	0.01	Primary schools rely more on multiple-choice items
<i>Comparison between schools in Bangkok (N = 39) and upcountry schools (N = 281)</i>			
% from final exam	0.14	0.01	Schools in Bangkok rely more on final exams
% from mid-term exam	-0.01	0.86	Not significant
% from continuous assessment	-0.07	0.21	Not significant
% from multiple-choice	0.01	0.89	Not significant

Table 2 shows that, for many assessment categories, different types of schools are similar in their assessment practices. For the differences, secondary schools in Bangkok make the most use of final exams, perhaps because these are the schools where the heaviest emphasis is placed on preparing students for ONET exams. Primary schools give greater weight to mid-term exams and multiple-choice items, findings for which the reasons are unclear.

Overall, the survey findings suggest that assessment practices likely to have negative washback are prevalent, but that all teachers use some open-ended continuous assessment. It does not appear to be the case that teachers disregard progressive assessment; rather, it is more

likely that there are factors pressuring teachers to use certain types of assessment leading to an overall imbalance in assessment methods. To find out what such factors might be, and to find out why, say, primary school teachers rely on multiple-choice, we need to investigate the reasons underlying teachers' decisions about assessment, the second goal of this study.

4. Why Do Teachers Use These Assessment Practices? (Study 2)

The assessment systems for ELT courses at Thai schools are not fixed, theoretically giving teachers a substantial amount of control in deciding how students are assessed on a given course. The survey findings show that teachers use a mix of assessment methods but also that the assessment mix is dominated by methods likely to have negative washback impacts. If teachers have control over assessment practices, why are so many potentially negative assessment methods being used? We might posit three possible reasons. First, teachers might not be aware of the consequences of their choices. Second, other potential benefits accruing to the methods, such as a need to show high levels of reliability, may outweigh the washback effects. Third, teachers may actually be constrained in what assessment methods they can use. In this study, we will focus on the third of these reasons, but our findings will also shed light on the other two reasons.

Teaching has been described as applied decision making, with decision making being the most basic teaching skill underpinning all others (Shavelson, 1973). Shavelson's theory has been applied extensively to explain teachers' classroom practices (e.g., Smith, 1996) but has rarely been applied to assessment practices. A key exception is McMillan and Nash's (2000) model of teacher decision making in assessment and grading practices. "The main tenet of the model holds that there was a tension between teachers' internal beliefs and values, and the realities of the classroom and other external factors imposed on them" (p. 6). The fact that teachers use some potentially beneficial assessment practices (which might match their beliefs), but rely more on practices with potentially negative washback (which might be imposed externally) suggests that this model may apply in the context of ELT assessment at Thai schools. Some support for this view is provided by Narathakoon et al. (2020) who found that teachers' assessment decisions were influenced by contextual factors, including educational policy, excessive workload, and time constraints.

To find out if McMillan and Nash's model applies, we gathered data on teachers' decision making and the reasons behind their choices of assessment practices. Since the reasons for any individual teacher are not easy to predict, we used open-ended data collection methods to allow teachers to express their own reasons. We therefore decided to use semi-structured interviews to investigate the teachers' decisions underpinning their use of assessment practices.

4.1 Methodology (Study 2)

To allow teachers freedom in their responses, a semi-structured interview protocol was constructed. This was organized around four topics to be covered in the interviews, namely, the teachers' beliefs about assessment, the reasons for their choices in assessment practices, constraints and resources affecting their choices, and descriptions of and reasons for an innovative assessment activity. (This last topic concerns the third goal of this research which is covered in

Study 3). Each topic had suggested sub-topics, and interviewers were free to cover the topic in whatever sequence felt natural in the context of a particular interview and to add further topics and sub-topics as appropriate. Interviewer training in active listening techniques for interviews was conducted (see Louw et al., 2011). The overall goal of the interviews was to encourage the teachers to express themselves as freely as possible within the general area of assessment decisions.

From the survey responses in Study 1, 29 teachers who were representative of the English language teacher population in Thailand were chosen for interviews. This cohort comprised both primary and secondary school teachers from around the country. Interviews were conducted face-to-face in Thai (to allow teachers to fully express themselves) and were recorded. Each interview lasted 30 minutes on average giving nearly 15 hours of data, which, when transcribed, comprised about 120,000 words.

Given the size of the data set, a corpus analysis approach was used. Using AntConc v3.5.8 (Anthony, 2019) as the corpus tool, word frequency analysis, collocation analyses and ngram analyses were conducted to identify key words and phrases in the data which were likely to be indicative of themes. Concordance lines for these words and phrases were checked to confirm the meanings and uses of these key words and phrases. Finally, stretches of text containing words with similar meanings were grouped into themes.

Two categories of themes were identified from the data: assessment practices and decision-making factors. The themes in the first category included proportions of scores assigned to different broad assessment types, task formats used, types of continuous assessment used, beliefs about purposes of assessment, and stakeholders involved in assessment. In this paper, we will focus on the first three of these assessment practice themes as the ones most likely to be related to washback effects. We will also distinguish assessment practices likely to create positive washback from those likely to create negative washback. The second category of theme that was identified from the data concerns factors which influence decision making, such as Ministry regulations, the need to prepare students for ONET exams, the desire to help students develop and be motivated, teacher workload, and students' future real-world needs.

The corpus was coded for the two types of themes separately so that co-occurrence of themes from the different types could be identified. This allows us to link the assessment issue themes with the decision-making factor themes to identify the reasons underlying teachers' decisions about assessment that could have positive or negative washback effects.

4.2 Results (Study 2)

The most frequent co-occurrences of assessment practice themes and decision-making factor themes were identified and these are shown in Table 3. The keywords most associated with each match between an assessment practice theme and a decision-making factor theme are shown together with their frequencies.

TABLE 3

Co-occurrences of Assessment Practices and Decision-making Factors

<i>Assessment issues</i>	<i>Ministry regulations</i>	<i>Preparing students for ONET</i>	<i>Student development and motivation</i>	<i>Teacher workload</i>	<i>Students' real-world needs</i>
Overall proportions of scores	curriculum (หลักสูตร) (58) assign (กำหนด) (85) final (ปลายภาค) (42) score (คะแนน) (39) proportion (สัดส่วน) (24)				
Use of closed task formats	choice (ข้อยส์) (114) multiple (มัลติเพิล) (49) subjective (อัตนัย) (14)	exam (ข้อสอบ) (386) choice (ข้อยส์) (114) multiple (มัลติเพิล) (49)		exam (ข้อสอบ) (386) score (คะแนน) (39) convenient (สะดวก) (13)	
Use of open task formats	subjective (อัตนัย) (14)		write (เขียน) (450) subjective (อัตนัย) (14)		
Use of continuous assessment (CA)	criterion (ตัวชี้วัด) (65) score (คะแนน) (39)		scores from CA คะแนนเก็บ (69) develop (พัฒนา) (44)		portfolio (พอร์ตโฟลิโอ) (37)

From Table 3, we can see that the factor most frequently mentioned as influencing teachers' decision making is Ministry regulations. These include the National Education Act, the Basic Education Curriculum, specific Ministry policies, and local policies. The frequency of mention of Ministry regulations implies that Thai education is controlled in a top-down fashion. These regulations appear to lack consistency in that they promote both continuous assessment and closed exams. We have already seen that the National Education Act and the Basic Education Curriculum promote a mix of assessment practices with an emphasis on continuous assessment. The most recent large-scale reform by the Ministry includes the Excellence Individual Development Plan which aims to reduce the number of exams for students and replace them with portfolios (Mala, 2020). On the other hand, in the interviews the teachers stated that the Ministry promoted the use of exams: "The Office of the Basic Education Commission assigns the teachers teaching Mathayom 3 and 6 [Grades 9 and 12] to make use of ONET as a guideline for their tests" (Teacher 12). A key issue here is that the number of regulations and requirements issued by the Ministry of Education is enormous. For instance, Watson Todd (2015) lists 11 initiatives by the Ministry in a three-month period that teachers were expected to follow, including promoting critical thinking, using portfolios, and preparing students for national tests. Many of these initiatives are contradictory, a point teachers are aware of (e.g., "You asked us to do Communicative Language Teaching focusing on communication but the tests you designed are not for communication at all" (Teacher 18)), but they still appear to influence teachers' decision making about assessment.

In the interviews, many of the teachers identified a shared set of overriding purposes guiding their decision making. Clearly dedicated to doing the best for their students, the teachers generally saw their key responsibilities as first, to help the students to develop and to motivate them in their learning; second, to prepare the students for their future real-world uses of English; and third, to prepare the students for the national exams such as ONET.

The first two of these responsibilities encourage teachers to use continuous assessment with positive washback impacts on their teaching as the following quotations show:

"By doing this activity, students will get participating marks. Even Mathayom 3 [Grade 9] students like this sort of activity, they have a chance to walk and talk to their classmates. They enjoy participating in the activities. If we only stand in front of the class and teach, the students will always feel sleepy." (Teacher 10)

"In terms of continuous assessment, it is not only end-unit quizzes, but also other kinds of work and reports. We do this to help students, and sometimes they ask their friends to help them too. We need to think back and ask ourselves about our teaching objectives first – do we want to help students really communicate in English or just want them to pass the course and be able to do nothing?" (Teacher 16)

The third responsibility of preparing students for national exams appears to be perceived as a necessary evil that results in assessment decisions with negative washback impacts on teaching. The focus on ONET exams forces teachers to change both what and how they teach:

"In the past, we faced big problems about ONET in terms of the content and the constructs of the test. When we planned to teach, we didn't care about ONET that much. We followed our lesson plans and focused on the learning outcomes. Therefore, we

required students to perform real-world tasks when we wanted to assess their abilities. However, what we did doesn't match with ONET. We therefore had to change our teaching strategies, and so we designed a test whose task formats correspond with ONET from the Mathayom 1 level [Grade 7]." (Teacher 12)

"What is in the test is clearly against OBEC regulations which aim to promote students' communication skills. The test is full of grammar items and any mistake is penalized, even mistakes such as adding -ed or -s after the verb. They [regulations and tests] are not similar at all. And then, teachers are blamed since they keep teaching and focusing on grammar rules. So look at the tests, they are different to what teachers want to teach; it's a huge conflict." (Teacher 19)

Despite their desire to do their best for students (even if this involved using grammar tests with negative washback), the teachers were constrained by an excessive workload. With numerous administrative responsibilities in addition to teaching, teachers find that they do not have the time to either design or mark the kinds of assessment activities they would like to use:

"There are high expectations that the students will be able to create some written work. But the thing is teachers teaching at high schools usually skip this kind of work, because the teachers need to check a lot of homework, so they refuse this kind of work. Only crazy teachers would do that. The normal ones normally say 'no'." (Teacher 4)

We posited three possible reasons why teachers may use potentially negative assessment methods. The interview findings show that the first of these, that teachers might not be aware of the consequences of their choices, does not apply. Rather, teachers are constrained in their choices of assessment methods by Ministry regulations and excessive workload, and they are forced to use some methods with potentially negative washback impacts so that they can achieve the goal of preparing students for national exams. These findings suggest that McMillan and Nash's (2000) model of assessment and grading practices decision making applies to English language teachers at Thai schools. In deciding what assessment practices they should use, teachers are juggling numerous conflicting factors including their own beliefs about best practices, Ministry requirements and potential constraints. While teachers want to use assessment practices with positive washback impacts, the reality of assessment at Thai schools is less than ideal.

5. How can positive assessment practices Be Promoted? (Study 3)

Although the assessment practices used in Thai schools are less than ideal, we saw that all teachers use some form of continuous assessment and that teachers are dedicated to doing their best for students. The key question we need to consider is how we can encourage teachers to use more assessment practices with potential positive washback to improve ELT at Thai schools. In one way, doing this should be fairly straightforward as it appears that such practices match the teachers' beliefs. The common problem of innovations not matching the teachers' beliefs (e.g., Carless, 2003) should therefore not be an issue in this context. What is preventing the teachers from using positive assessment practices are the situational constraints, and these need to be addressed. In the research report written for the project described in this article which was submitted to the Ministry of Education (Watson Todd et al., 2020a), we included recommendations addressing the

constraints. For example, one recommendation stated “The National Institute of Educational Testing Services should consider altering the format of national exams such as ONET so that they do not rely exclusively on multiple-choice testing”. Another stated “When promulgating new regulations or guidelines, authorities should consider the potential impact of the new regulations on assessment practices, and, where possible, regulations should be designed to promote assessment practices with positive washback”. Addressing these factors relies on interventions by the authorities and is beyond the control of teachers.

The one factor that can be addressed is workload. Although the number of responsibilities assigned to any teacher is a school or Ministry decision, resources can be provided that allow teachers to achieve their goals without needing to devote an excessive amount of time. We saw that excessive workload prevented some teachers from designing the assessment activities they wanted to use, an issue which providing appropriate resources may alleviate. A potentially valuable approach, then, is to create a database of freely available relevant assessment activities that teachers can download and use, allowing them to gain the benefits of the assessment activities without requiring them to devote time to designing the activities. If we create such a database, the activities offered must have two essential characteristics: they must be relevant and they must be beneficial.

The clearest way to ensure that the assessment activities are relevant is to offer activities that Thai school teachers designed themselves, since these are the activities most likely to match the specific context. In the survey described in Study 1, one section asked teachers to give the details of one assessment activity that they were proud of. 303 of the survey respondents answered this section of the questionnaire giving us a large pool of potential activities to choose from. The key issue, then, is how we can judge which of these activities are likely to be beneficial.

The differential effects of various assessment practices have been examined in previous research with the vast majority of studies suggesting that continuous assessment is more likely to allow progressive learning objectives to be reached than tests. For example, continuous assessment is more likely to promote meaningful, authentic and creative communication (Brown, 2016), to promote learner autonomy (Everhard, 2015), and to lead to greater proficiency gains (Ross, 2005). These findings imply that we should focus on those activities suggested by teachers which can be considered continuous assessment, an implication that is not very useful since the vast majority of the 303 assessment activities submitted by the teachers in the survey are continuous assessment activities. With no clear guidelines in the literature beyond the potential benefits of continuous assessment as a whole, we needed to identify the aspects of assessment activities that make them more likely to be beneficial.

5.1 Methodology (Study 3)

To identify the characteristics associated with potentially beneficial assessment activities, we used two sources of data. First, we examined the interviews with the teachers. One part of the interviews concerned the teachers’ reasons for choosing the activity they had submitted in the survey. The relevant responses in the interviews were analyzed for recurring mentions of particular characteristics (a frequency measure) and for insightful quotations (a saliency measure).

Second, using a representative sample of 18 activities submitted by teachers in the survey, eight judges with expertise in materials or test design were asked to rank these activities into six

groups (i.e., the ‘best’ three activities, the next ‘best’ three activities etc.). After ranking the activities, they were asked to give brief justifications for their higher rankings. These justifications were analyzed for frequency of concepts indicative of the characteristics of preferred activities.

5.2 Results (Study 3)

In the interviews with teachers, the adjective most frequently associated with ideal assessment activities is *real* most usually in the phrase *real-world use* implying that activities must be practical. The next most frequently used adjectives are *fun* and *varied* (the latter matching the recommendations of the National Education Act). The need for assessment activities to be fun reflects the teachers’ awareness that assessment practices have affective impacts on the students and these need to be considered in addition to the traditional concerns of reliability and validity:

“For assessment and evaluation, if it concerns language, what assessment method should we use to help learners? First, personally, attitude needs to be improved. The first priority is the attitude. Positive attitude towards English subjects needs to be improved.”
(Teacher 4)

Such concerns with the affective impacts of assessment activities do not mean that teachers are not concerned with validity. One reason that teachers favor the use of continuous assessment is that it has content validity that exams lack:

“For our purpose, we want students to communicate, to be able to communicate in English. However, the exam does not focus on communication, but on grammatical accuracy. For communication, if they talk with foreign students, they must be able to speak. But, the test does not include any part which students are interviewed or are engaged in a conversation task that will allow them to use their communication skills.”
(Teacher 18)

The judges’ rankings of the 18 representative assessment activities were relatively consistent. For example, the highest ranked activity was ranked in the top group by 7 of the 8 judges, and the lowest ranked activity was ranked in the lowest group by 7 of the 8 judges. Half of the activities had a standard deviation of ranking of less than 1 (suggesting very high levels of agreement between judges), and only 4 of the activities had a standard deviation of ranking of more than 1.5 (with a maximum of 1.69). This level of consistency suggests that common criteria allowing us to identify quality assessment activities do exist.

Turning to the judges’ justifications of their rankings, the adjectives most frequently used to describe higher-ranked activities were *interesting* (5 of the 8 judges), *challenging* (5 judges), and *practical* (3 judges). In addition, 4 judges used phrases related to *authenticity* (e.g., “The teacher brings real-life issues to English class”) and related to *achieving curricular goals* (e.g., “serves the indicators in the curriculum”). Most importantly for our purposes, 4 judges mentioned points that can be related to positive washback impacts from the activities onto teaching (e.g., the students learn how to “organize their script ... consult with the teacher, ... and learn how to edit”).

5.3 Implementation

From the interviews and the rankings, the characteristics associated with beneficial assessment activities are 1. Practicality; 2. Interest and fun; 3. Validity in meeting curricular goals, especially communication; 4. Challenge; and 5. Authenticity. Using these characteristics, we evaluated the 303 assessment activities submitted by the teachers in the survey to identify the most beneficial activities that could be shared so that students throughout Thailand could benefit without having to devote substantial amounts of time to develop their own activities.

Looking through the submitted activities, some could not be considered as the description provided was too perfunctory for a judgment about the quality to be made, while a few did not use continuous assessment which, as we have seen, is most likely to be beneficial. The remaining activities were informally rated for the extent to which they exhibited the desired characteristics, until we had a small pool of activities remaining. From this, we created a database of 39 activities which covered both primary and secondary education and which met a range of learning objectives.

For each of the activities in this database, the teacher who had originally submitted the activity was contacted for permission to use the activity and to elicit more details. The full details of the activity were then redesigned into an attractive format for dissemination.

To illustrate the progression from original submission in the survey to final version for dissemination, we will use two activities as examples. Table 4 shows the original information about these activities which was submitted in the survey.

TABLE 4
Information about Two Activities Submitted in the Survey

Type of information	Activity 1	Activity 2
<i>Name of activity of which you are proud</i>	Food and Drink	Mobile phone user survey
<i>Materials needed for the activity</i>	1. Teacher gives an example of making a particular food or drink 2. Students bring their own favorite recipes from home	Worksheet Images PPL
<i>Assigning marks</i>	1. Self-assessment 2. Peer assessment 3. Teacher assessment	Report: 5 points Presentation: 10 points
<i>Activity details</i>	1. Divide students into groups of 4-5 2. Each group of students works together on a recipe for their favorite dish, sequencing the steps in easy-to-understand English 3. Each group presents a simulation of the preparation and cooking process in front of class 4. Students' presentations are assessed	1. Read the article entitled 'Are cell phones safe?' as a reading circle. Complete the vocabulary and comprehension activities 2. Gain an understanding of the 'Mini Survey on Mobile Phone User Behavior' and practice using the survey which consists of 12 questions on 5 issues. Students work in groups of 3 (mixed-ability) 3. Spend 2 weeks outside the classroom conducting interviews of other students

		based on the survey and bring the data to class to calculate means as a summary
		4. Learn how to give presentations; practice the presentations; give the presentations (7 minutes per group)
		5. Peers and the teacher give feedback
<i>Learning objectives</i>	Being able to explain the sequence of steps in cooking	Interviewing and presentation skills

From Table 4, we can see that different teachers had slightly different understandings of the nature of the information about the activities requested. Nevertheless, with more detailed information collected from the teachers directly, a standard format for the activities could be constructed. For each activity, 11 categories of information were presented:

1. Summary of activity
2. Language skill(s) (e.g., speaking, listening)
3. Language function (e.g., asking and answering personal questions)
4. Type of activity (e.g., role-play, presentation)
5. Class organization (e.g., pairwork, individual work)
6. Time for the activity
7. Learning objectives
8. Materials required
9. Stages in running the activity
10. Performance evaluator (e.g., teacher, peers)
11. Scoring criteria

This information was then designed in attractive formats for dissemination. The final disseminated versions of the activities given in Table 4 are shown in Figures 1 and 2.

Figure 1 shows an assessment activity created by Ajarn Wasana Kamhuayhan from Chiang Rai. The activity aims to assess students' speaking skills when giving a cooking demonstration. Students in Pathom 4-6 (Grades 4-6) work together in groups of 4 or 5. Using real ingredients, they prepare a dish they have chosen. While doing this, the students explain the instructions as in a cooking demonstration. They are marked for the appropriacy of their explanation and for their pronunciation (5 marks each) with both their peers and the teacher scoring their presentation.

Figure 2 shows an assessment activity created by Ajarn Rattapol Srichaiwong from Chiang Mai. Designed for students in Mathayom 4 to 6 (Grades 10-12), the activity aims to assess students' oral presentation through a 4-skills integrated task. Working in pairs or small groups, students prepare and conduct an interview survey on mobile phone user behavior and present their findings. The students work step-by-step through a series of assigned activities, such as reading a text about cell phones, summarising the text, practicing pronunciation, and developing interview questions. After analyzing their interview data, students need to write a short report and give a presentation. During the oral presentations, the students' peers act as commentators, while the teacher's role is to give constructive feedback. Students are graded for both their report (5 marks) and their presentation (10 marks). Their presentations are marked for five criteria, namely, vocabulary, grammatical accuracy, fluency, comprehensibility and effort to communicate.

The final versions of the 39 selected assessment activities were disseminated in two formats. First, a book (Watson Todd et al., 2020b) containing all of the activities as photocopiable sheets plus guidance on how to use them was published and distributed to 1,000 schools throughout Thailand. Second, a website containing downloadable versions of the activities was created (<http://sola.pr.kmutt.ac.th/assessments/>). The website allows the database to be searched by education level, language skill and class organization. Within the first 80 days after the website was uploaded, there were over 22,000 visits from over 2,800 individual visitors, suggesting that beneficial assessment activities have now become available for teachers to use with potential positive impacts on ELT in Thailand.

FOOD AND DRINK

(อาหารและเครื่องดื่ม)

ระดับชั้น ป. 4-6

กิจกรรมโดยย่อ

การพูดบรรยายขั้นตอนการทำอาหารโดยผ่านกิจกรรมการสาธิตการทำอาหารจากวัตถุดิบจริง

ทักษะทางภาษา

การพูดเพื่อการนำเสนอ

หน้าที่ทางภาษา

บรรยายขั้นตอนการทำอาหาร

ลักษณะของกิจกรรม

การสาธิตทำอาหาร

ประเภทกิจกรรม

งานกลุ่ม

เวลาในการทำกิจกรรม

1 คาบเรียน

จุดประสงค์การเรียนรู้

นักเรียนสามารถบอกลำดับขั้นตอนของการทำอาหาร

สื่อ/อุปกรณ์ประกอบ

- โปรแกรมการนำเสนอ
- วัตถุดิบในการทำอาหารหรือเครื่องดื่มที่นักเรียนนำมาจากบ้าน

ขั้นตอนการจัดกิจกรรม

- แบ่งนักเรียนออกเป็นกลุ่ม กลุ่มละ 4-5 คน
- ให้นักเรียนแต่ละกลุ่มร่วมกันคิดเมนูอาหารที่ชื่นชอบและลำดับขั้นตอนการทำอาหารหรือเครื่องดื่มนั้นๆ
- นักเรียนแต่ละกลุ่มนำเสนอขั้นตอนการทำอาหารหน้าชั้นเรียนโดยจำลองสถานการณ์เป็นห้องครัวรายการทำอาหาร
- ให้เพื่อนในห้อง และครู รวมถึงตนเองประเมินให้คะแนน โดยจะมีการอธิบายเกณฑ์ให้คะแนนก่อน

ผู้ประเมิน

ครู/เพื่อน

เกณฑ์การให้คะแนน

เกณฑ์	คะแนน
การบรรยายถูกต้องตามขั้นตอน	5คะแนน
การออกเสียง	5คะแนน
รวม	10คะแนน

Credit: อ. วาสนา คำช่วยหาญ โรงเรียนบ้านสันโค้ง (เขียงราษฎร์ราษฎร์) จ. เชียงราย
happyoun@hotmail.com

FIGURE 1. Disseminated Version of Food and Drink Assessment Activity

Mini Survey on Mobile Phone



User Behavior

(สำรวจพฤติกรรมผู้ใช้โทรศัพท์มือถือ)



ระดับชั้น ม.4-6

กิจกรรมโดยย่อ

การใช้ภาษาอังกฤษสัมภาษณ์ Mobile Phone Users แล้วนำข้อมูลมาวิเคราะห์ หาค่า และนำเสนอได้

ทักษะทางภาษา

การฟัง การพูด การอ่าน และการเขียน

หน้าที่ทางภาษา

ใช้ภาษาอังกฤษเพื่อการขอข้อมูลและนำเสนอ

ลักษณะของกิจกรรม

Interaction patterns ปฏิสัมพันธ์ระหว่างนักเรียนกับนักเรียน ในขณะที่ทำกิจกรรม

โดยใช้รูปแบบปฏิสัมพันธ์กลุ่มใหญ่ (big groups) รูปแบบปฏิสัมพันธ์กลุ่มเล็ก (small groups)

รูปแบบปฏิสัมพันธ์คู่ (pairs) รูปแบบปฏิสัมพันธ์ผสมหรือคละเคล้า (mingle))

ประเภทกิจกรรม

งานคู่และกลุ่ม

เวลาในการทำกิจกรรม

8 คาบ คาบละ 50 นาที

สื่อ/อุปกรณ์ประกอบ

- 1.รูปภาพ
- 2.บทอ่านเรื่อง Cell Phones: Hang Up or Keep Talking?
- 3.แบบสำรวจ Mini Survey on Mobile Phone User Behavior
- 4.ใบความรู้ Useful Language for oral presentation
- 5.แบบสรุปการเขียนรายงานแบบสำรวจ

จุดประสงค์การเรียนรู้

- 1.สามารถสรุปใจความหลักและเชื่อมโยงเนื้อหาในบทอ่าน Mobile Phone Users
- 2.สามารถใช้ภาษาอังกฤษสัมภาษณ์รุ่นน้องในโรงเรียนจำนวน 4 คน
- 3.สามารถนำข้อมูลมาวิเคราะห์ หาค่า และนำเสนอได้

ขั้นตอนการจัดกิจกรรม

- 1.เรียนบทอ่านเรื่อง Cell Phones: Hang Up or Keep Talking? ทีละย่อหน้า โดยเริ่มจากเรียนคำศัพท์ยากโดยใช้เทคนิคจำทีละ 7 (seven at a time)
- 2.ทำกิจกรรมการอ่านออกเสียงโดยใช้เทคนิค circle reading และทำความเข้าใจในบทอ่าน
- 3.นักเรียนและครูช่วยกันสรุปใจความสำคัญให้ถูกต้องและเหมาะสม
- 4.นักเรียนทำงานกลุ่ม ๆ ละ 3-4 คน
- 5.เรียนคำถาม Mobile Phone User Behavior Questions จำนวน 12 ข้อ 5 ประเด็น (purpose, connection, time & duration, expense, do it or not) และฝึกใบขึ้นเรียนก่อนโดยใช้เทคนิค mingle
- 6.แบ่งกลุ่มนักเรียน กลุ่ม ๆ ละ 3 คน (เก่ง อ่อน ปานกลาง) ใช้เวลาว่างในการสัมภาษณ์ 2 สัปดาห์ โดยนักเรียน 1 คน สัมภาษณ์รุ่นน้องจำนวน 4 คน และเติมข้อมูลลงในใบงานพร้อมถ่ายภาพขณะทำกิจกรรม
- 7.นำข้อมูลที่ได้อาจจัดเรียงกลุ่มข้อมูล วิเคราะห์ หาค่าเฉลี่ย และเขียนรายงานและนำเสนอ PowerPoint
- 8.เรียนวิธีการนำเสนอ ฝึกการนำเสนอภายในกลุ่ม และฝึกการนำเสนอหน้าห้องโดยการจับเวลา กลุ่มละไม่เกิน 7 นาที
- 9.เพื่อนทำหน้าที่เป็น commentator และครูให้ข้อมูลย้อนกลับ (praise and give feedback)

Credit: อ.รัฐพล ศรีขัยวงศ์ โรงเรียนเวียงแหงวิทยาคม จ.เชียงใหม่
Email: aue_rattapol@hotmail.com

Mini Survey on Mobile Phone User Behavior

(สำรวจพฤติกรรมผู้ใช้โทรศัพท์มือถือ)

ระดับชั้น ม.4-6



ผู้ประเมิน

เพื่อน และครู

เกณฑ์การให้คะแนน

เต็ม 15 คะแนน



เกณฑ์	คะแนน
รายงาน	5 คะแนน
การพูดนำเสนอ 5 ประเด็น	10 คะแนน

เกณฑ์การประเมินการพูดนำเสนอ 5 ประเด็น	ระดับคะแนน			
	1	2	3	4
1. การใช้คำศัพท์ (vocabulary)				
2. ความถูกต้องทางไวยากรณ์ (grammatical accuracy)				
3. ความคล่องแคล่ว (fluency)				
4. ความสามารถพูดให้ผู้อื่นเข้าใจ (comprehensibility)				
5. ความพยายามในการสื่อสาร (effort to communicate)				




Credit: อ.รัฐพล ศรีชัยวงศ์ โรงเรียนเวียงแหงวิทยาคม จ.เชียงใหม่
Email: aue_rattapol@hotmail.com

FIGURE 2. Disseminated Version of Mobile Phone User Survey Assessment Activity

6. Conclusion

The low quality of education, especially English language education, in Thailand is a cause for concern. There is a clear need for innovations to improve the situation, but the history of innovations can be seen as a litany of failures with only a scattering of successes (Kirschner et al., 2004). For an innovation to have a chance of succeeding, it needs to be relevant, acceptable and feasible (Markee, 1997).

From the findings in Study 1, the fact that all teachers use some continuous assessment activities, although perhaps not enough to have major washback impacts, suggests that these activities are acceptable and feasible. From the interviews in Study 2, the factors that appear to be preventing teachers from using more such activities are regulations, the need to prepare students for ONET exams, and excessive workload. Preparing activities that teachers can use easily can reduce some workload concerns, especially if the activities are designed by teachers in similar contexts ensuring relevance. Study 3, then, made 39 teacher-designed assessment activities with potentially positive washback effects freely available to Thai school teachers. We believe that changing assessment practices at schools to include more activities likely to promote positive washback, such as progressive continuous assessment activities, has a good chance of leading to meaningful change in Thai education.

Acknowledgements

This research is part of a project funded by the National Research Council of Thailand entitled 'A survey of English language assessment practices at schools throughout Thailand focusing on practices with positive washback' (Thailand Research Fund Project RDG61H0028). We would also like to thank Ajarn Pongpet Pongsart and Ajarn Khanitta Piboonkanarax for their help with the project.

About the Authors

Richard Watson Todd: Associate Professor and Associate Dean for Research in the School of Liberal Arts, King Mongkut's University of Technology Thonburi (SoLA, KMUTT).

Duangjaichanok Pansa: a teacher in SoLA, KMUTT.

Natjiree Jaturapitakkul: Assistant Professor and Assistant Dean for Assessment and Evaluation in SoLA, KMUTT.

Nawiya Chanchula: a teacher in SoLA, KMUTT.

Punjaporn Pojanapunya: Assistant Professor and a researcher in SoLA, KMUTT.

Saowaluck Tepsuriwong: Assistant Professor and Associate Dean for Academic Affairs and Quality Assurance in SoLA, KMUTT.

Stuart G. Towns: a teacher in SoLA, KMUTT.

Wannapa Trakulkasemsuk: Associate Professor and Head of the Department of Language Studies, SoLA, KMUTT.

REFERENCES

- Anthony, L. (2019). *AntConc* (Version 3.5.8).
<https://www.laurenceanthony.net/software/antconc/>
- Basic Education Commission. (2008). *Basic education core curriculum BE 2551 (AD 2008)*. Ministry of Education.
- Brown, A. (2013). Uses of language assessments. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5979-5985). Wiley-Blackwell.
<https://doi.org/10.1002/9781405198431.wbeal1237>
- Brown, J. D. (2005). *Testing in language programs*. McGraw-Hill.
- Brown, J. D. (2016). Assessment in ELT: Theoretical options and sound pedagogical choices. In W. A. Renandya & H. P. Widodo (Eds.), *English language teaching today: Linking theory and practice* (pp. 67-82). Springer. https://doi.org/10.1007/978-3-319-38834-2_6
- Burke, K. (1999). *How to assess authentic learning*. Skylight Professional Development.
- Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, 31(4), 485-500. <https://doi.org/10.1016/j.system.2003.03.002>
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–46). John Wiley & Sons.
<https://doi.org/10.1002/9781118411360.wbcla071>
- English First. (2019). *EF English proficiency index*. Retrieved from
<https://www.ef.edu/~/media/centralefcom/epi/downloads/full-reports/v9/ef-epi-2019-english.pdf>
- ETS. (2020). *TOEFL iBT test and score data summary 2019*. Retrieved from
https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf
- Everhard, C. J. (2015). The assessment-autonomy relationship. In C. J. Everhard & L. Murphy (Eds.), *Assessment and autonomy in language learning* (pp. 8-34). Palgrave Macmillan.
https://doi.org/10.1057/9781137414380_2
- Fitzpatrick, D. (2011). *Making sense of the English language policy in Thailand: An exploration of teachers' practices and dispositions* (Doctoral thesis, University of Exeter, Exeter, UK). Retrieved from <https://ore.exeter.ac.uk/repository/handle/10036/3314>
- Foley, J. A. (2005). English in ... Thailand. *RELC Journal*, 36(2), 223-234.
<https://doi.org/10.1177/0033688205055578>
- IMD. (2020). *IMD world competitiveness yearbook 2020, digital 2020, talent 2019: Summaries: country profile: Thailand*. Retrieved from
<https://www.imd.org/globalassets/wcc/docs/wco/pdfs/countries-landing-page/th.pdf>
- Jianrattanapong, A. (2011). Positive washback from Thai university entrance exams. *Language Testing in Asia*, 1(1), 50-61. <https://doi.org/10.1186/2229-0443-1-1-50>
- Kirschner, P. A., Hendricks, M., Paas, F., Wopereis, I., & Cordewener, B. (2004). Determinants for failure and success of innovation projects: The road to sustainable educational innovation. *Proceedings of the Association for Educational Communications and Technology convention*, Chicago, IL, 19-24 October 2004 (pp. 359-368). Retrieved from
<https://files.eric.ed.gov/fulltext/ED485042.pdf>

- Louw, S., Watson Todd, R., & Jimarkon, P. (2011) Active listening in qualitative research interviews. *Selected proceedings of the international conference: Doing research in applied linguistics* (pp. 71-82). 21-22 April 2011, Bangkok: King Mongkut's University of Technology Thonburi.
- Mala, D. (2020, August 18). Major education overhaul in works. *Bangkok Post*. Retrieved from <https://www.bangkokpost.com/thailand/general/1969855/major-education-overhaul-in-works>
- Markee, N. (1997). *Managing curriculum innovation*. Cambridge University Press.
- McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. Retrieved from <https://files.eric.ed.gov/fulltext/ED447195.pdf>
- Narathakoon, A., Sapsirin, S., & Subphadoongchone, P. (2020). Beliefs and classroom assessment practices of English teachers in primary schools in Thailand. *International Journal of Instruction*, 13(3), 137-156.
- Nonthaisong, K. (2015). *The impact of the English language policy in the Basic Education Core Curriculum B.E. 2551 (A.D. 2008) on English instruction of Thai secondary EFL teachers: A multiple case study in northeast Thailand*. (Doctoral thesis, University of Alabama, Tuscaloosa, USA). Retrieved from https://ir.ua.edu/bitstream/handle/123456789/2417/file_1.pdf?sequence=1&isAllowed=y
- Noom-ura, S. (2013). English-teaching problems in Thailand and Thai teachers' professional development needs. *English Language Teaching*, 6(11), 139-147. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1078705.pdf>
- OECD. (2016). *Education in Thailand: An OECD-UNESCO perspective*. Paris: OECD Publishing. Retrieved from <https://www.oecd.org/publications/education-in-thailand-9789264259119-en.htm>
- Office of the National Education Commission. (1999). *National Education Act of B.E.2542 (1999)*. Office of the National Education Commission.
- Piboonkanarax, K. (2007). *A survey of secondary school evaluation procedures focusing on continuous assessment* (Unpublished Master's thesis). King Mongkut's University of Technology Thonburi, Bangkok, Thailand.
- Poonpon, K. (2012). Integrating English speaking tests in an EFL classroom: What teacher can learn from. Retrieved from https://www.iaea.info/documents/paper_4d32efbe.pdf
- Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language Testing*, 25(1), 127-143. <https://doi.org/10.1177/0265532207083748>
- Ross, S. J. (2005). The impact of assessment method on foreign language proficiency growth. *Applied Linguistics*, 26(3), 317-342. <https://doi.org/10.1093/applin/ami011>
- Shavelson, R. J. (1973). What is the basic teaching skill? *Journal of Teacher Education*, 24(2), 144-151. <https://doi.org/10.1177%2F002248717302400213>
- Smith, D. B. (1996). Teacher decision making in the adult ESL classroom. In D. Freeman & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp. 197-216). Cambridge University Press.
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29. <https://doi.org/10.1191/1362168805lr152oa>

- Thongsri, M., Charumanee, N., & Chatupote, M. (2006). The implementation of 2001 English language curriculum in government secondary schools in Songkhla. *ThaiTESOL Bulletin*, 19(1), 60-94.
- Tomlinson, B. (2005). Testing to learn: A personal view of language testing. *ELT Journal*, 59(1), 39-46. <https://doi.org/10.1093/elt/cci005>
- Watson Todd, R. (2008) The impact of evaluation on Thai ELT. *Selected proceedings of the 12th English in South–East Asia international conference: Trends and directions* (pp. 118 – 127). Bangkok: King Mongkut’s University of Technology Thonburi.
- Watson Todd, R. (2015) National-level educational innovations in Thailand. In Darasawang, P. & Reinders, H. (Eds.) *Innovation in language teaching and learning: The case of Thailand* (pp. 15-28). Palgrave Macmillan. https://doi.org/10.1057/9781137449757_2
- Watson Todd, R., Tepsuriwong, S., Trakulkasemsuk, W., Jaturapitakkul, N., Pojanapunya, P., & Chanchula, N. (2020a) โครงการการสำรวจแนวปฏิบัติด้านการวัดและประเมินผลรายวิชาภาษาอังกฤษที่ใช้ในโรงเรียนไทยในปัจจุบันโดยเน้นแนวปฏิบัติที่ส่งผลกระทบเชิงบวก [A survey of current English language assessment practices at schools throughout Thailand focusing on practices with positive washback]. Report submitted to the National Research Council of Thailand.
- Watson Todd, R., Tepsuriwong, S., Pojanapunya, P., Trakulkasemsuk, W., Jaturapitakkul, N., Pansa, D., Towns, S., Pongsart, P., Piboonkanarax, K., & Chanchula, N. (2020b) สุดยอดกิจกรรมการประเมินผลวิชาภาษาอังกฤษ (คู่มือครู) [The best activities for assessing English language]. Bangkok: School of Liberal Arts, King Mongkut’s University of Technology Thonburi.