

Language Testing

<http://ltj.sagepub.com/>

Convergent vs. divergent assessment: Impact on college EFL students' motivation and self-regulated learning strategies

Shu-Chen Huang

Language Testing 2011 28: 251 originally published online 22 February 2011

DOI: 10.1177/0265532210392199

The online version of this article can be found at:

<http://ltj.sagepub.com/content/28/2/251>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language Testing* can be found at:

Email Alerts: <http://ltj.sagepub.com/cgi/alerts>

Subscriptions: <http://ltj.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ltj.sagepub.com/content/28/2/251.refs.html>

>> [Version of Record](#) - Apr 14, 2011

[OnlineFirst Version of Record](#) - Feb 22, 2011

[What is This?](#)

Convergent vs. divergent assessment: Impact on college EFL students' motivation and self-regulated learning strategies

Shu-Chen Huang

National Chengchi University, Taiwan

Language Testing

28(2) 251–271

© The Author(s) 2010

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0265532210392199

<http://ltj.sagepub.com>



Abstract

This study examined two types of classroom assessment events, the more closed convergent assessments (CA) versus the more open-ended divergent assessments (DA), to see if they influence learners differently in terms of motivation and learning strategies. Participants were 105 college freshmen in Taiwan with the same instructor placed under one listening and two speaking classes. In these intact groups, each student experienced two types of assessment, a more traditional test (the CA) and a group performance assessment (the DA). Immediately after they experienced each assessment event, participants reported on their task-specific motivation and learning strategies. Results indicated that CAs were better received by high self-efficacious students and DAs by low scorers. After controlling for self-efficacy, a comparison revealed that students reacted differently in the listening and speaking classes. More specifically, in speaking student motivation and strategy were higher for the DA than for the CA, whereas in the listening class the pattern was reversed. Local educational culture and the nature of assessment events are then discussed.

Keywords

assessment for learning, classroom assessment, motivation, self-regulation

Researchers and practitioners have long been advocating the creation of assessments other than decontextualized, norm-referenced measurements. While the potential impact of such alternative assessments as a component of the complete learning cycle is now generally acknowledged, research in this field is still in its infancy. About a decade ago, McNamara (2001), in rethinking the priorities and responsibilities of language testing research, discussed the competing demands of validity, management, and teacher/learner needs, especially on classroom-based assessments. Among them, according to McNamara, the demands of teachers and learners appeared to be least favorably served by current

Corresponding author:

Shu-Chen Huang, 64, Sec. 2, Zhi-nan Road, Taipei, Taiwan, Republic of China 116.

Email: huang91@nccu.edu.tw

research. This awareness of insufficiency has gradually become a central concern in the literature, and a renewed interest in formative teacher assessment now offers an opportunity to reintegrate teaching, learning, and assessment (Leung, 2004), as well as aligning tests, curricula, and teaching (Cumming, 2009). There are also calls for tests to 'bias for the best' (Fox, 2004) or to 'blur the lines between teaching and testing' (Connor-Greene, 2000). Similar attempts and developments in neighboring fields, such as the teaching of science and mathematics, have been systematically reported (Harlen & Winter, 2004); many have been based on the work of the Assessment Reform Group (Black, Harrison, Lee, Marshall, & Wiliam, 2003). Alternative assessment concepts have also been extensively discussed, including performance assessment (Adair-Hauck, Glisan, Koda, Swender, & Sandroch, 2006), formative assessment (Rea-Dickins & Gardner, 2000), dynamic assessment (Leung, 2007; Poehner, 2007; Poehner & Lantolf, 2005), assessment *for* learning as opposed to assessment *of* learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black & Wiliam, 1998; Chappuis & Stiggins, 2002), and learning-oriented assessment (Carless, 2007). Large-scale policy and educational reform, involving teachers doing assessments in their own classrooms, have already taken place in China (Xu & Liu, 2009), Hong Kong (Hamp-Lyons, 2009), the United States (Adair-Hauck et al., 2006), the United Kingdom (Black et al., 2003), and many other parts of the world.

The aforementioned accounts notwithstanding, there remain a lot of unknowns in alternative forms of assessment. First and foremost, reliance on traditional psychometric testing theories has made it difficult to identify a research agenda, synthesize studies, and accumulate knowledge over time and contexts, partly due to a lack of theorization in teacher classroom assessment (Davison & Leung, 2009; Lantolf & Poehner, 2004; Poehner, 2007; Teasdale & Leung, 2000). As attested by both Teasdale and Leung (2000) and Torrance and Pryor (2001), many of our expectations for formative assessment and its potential benefits for learning are not well founded (unlike the negative consequences brought about by large-scale testing, which are relatively well understood). Many assumptions about the effectiveness of formative assessments amount to speculation, and more empirical research from real classrooms is needed.

Researchers have recently started to investigate teacher assessments in real classrooms. Among the empirical studies that have been completed, many focus on the role of teachers. For example, Butler (2009) found that teachers observe and assess students' foreign language performance very differently within and across school levels. Butler concluded that the observed variation had to do with different teaching contexts. Xu and Liu (2009) identified three structural conditions that affect teacher assessment knowledge, including past experiences, power distance in the workplace, and the specific contexts the teacher is situated in. Adair-Hauck et al. (2006) discovered that a large formative assessment program had a positive effect on teacher perceptions and instructional practices. These recent findings have provided data from real teaching/learning contexts and are valuable for the construction and refinement of classroom assessment theories.

One common theme emerging from these studies is that past experiences, and the specific social milieu and educational contexts, are important factors influencing teacher-based assessments. Studies on teachers, however, tell just part of the story. In an assessment movement that claims learning as its central concern (Black et al., 2003), how learners perceive and react to these assessment alternatives is critical information for

teachers and researchers. Such information should not be assumed as being based on studies of teachers.

To what degree would student perceptions and reactions matter? The answer is illustrated well in a study on the authenticity of an assessment task. Spence-Brown (2001) examined a teaching and assessment activity in a college Japanese language course, in which students were asked to interview native speakers outside the classroom. While the original intention of the instructor was to create authentic communication opportunities, the assessment dimension, even though the stakes were low in that it only accounted for 10% of the final grade, resulted in various learner strategies that eventually made the student experience inauthentic. Instances found included, among others, rehearsing and re-taping parts of the interview, saying what was easiest to say correctly rather than what students really wanted to say, and disguising comprehension difficulties by feigning understanding. The author argued that the authenticity of an assessment activity cannot be viewed solely from its design and from teachers' perspectives. It is its implementation (i.e. how learners interpret and react to it) that makes the assessment task what it is. Without this understanding of the role of the learner, teachers are very likely to interpret student performance inaccurately and go astray when feedback needs to be provided for the follow-up learning cycles.

When so much is involved in what learners think and do in responding to assessment tasks, what specifically should we look for in understanding whether our classroom assessments have done what we intended? Two aspects, although hardly ever associated with traditional testing and assessment, are of critical importance when it comes to assessments *for* learning rather than assessments *of* learning; they are learning motivation and strategy use. Positive motivational beliefs may lead to deep-processing strategies, which in turn enable students to learn better (Duncan & McKeachie, 2005; Pintrich, 2003). In addition, a meta-analysis indicates that existing empirical classroom assessment studies are found to be framed much more under motivational and cognitive theories than measurement frameworks (Brookhart, 1997). The significant and relevant constructs of motivation and learning strategy are discussed below.

Foreign language motivation and learning strategies

The study of foreign/second language (L2) learning motivation can be traced back to Gardner's social psychological model (1985), developed in bilingual Canada. Its most widely known concepts are integrative and instrumental orientation, the former being the 'willingness to be like valued members of the language community,' and the latter 'pertaining to the potential pragmatic gains of L2 proficiency' (Dörnyei, 2001, p. 49). They help explain why some learners expend more effort and persist longer in L2 learning than others. Until the 1990s, concerns for depicting and explaining L2 motivation in formal learning situations, accompanied by the developments of motivation theories in educational psychology, brought about an 'educational shift' (Dörnyei, 2001, p. 104) in the field. At that time, one of the more educationally friendly constructs, developed from empirical results, was Dörnyei's tripartite framework (Dörnyei, 1994). It delineated L2 motivation from the more macro to the more micro level in three dimensions, namely the social dimension, the personal dimension, and the learning situation dimension. Again,

integrative and instrumental motivation are the major constituents in the social dimension related to the specific L2 being learned. When similar investigations were carried out in Taiwan, where the pursuit of academic degrees and high test scores is highly valued, one additional motivational cluster, requirement motivation, was identified (Chen, Warden, & Chang, 2005; Warden & Lin, 2000). According to these researchers, it is the requirement set by educational institutions rather than integration or instrumentation that underlies most students' enrollment in L2 courses. They cautioned that if we import theories developed in other social contexts and fail to take into consideration cultural influences, our learners may be led in the wrong direction by motivational strategies.

Moving from the social dimension, Dörnyei (1994) points out in the personal dimension the significance of individual learner characteristics, including learners' self-confidence and need for achievement. For self-confidence, the concept of self-efficacy (Bandura, 1977) has been at center stage. According to Bandura, self-efficacy refers to personal judgment of one's capabilities to organize and execute courses of action to attain designated types of educational performances. It is domain specific, that is, it varies in the same person by subject matter and situation. A number of empirical studies and meta-analyses have consistently demonstrated that a high level of self-efficacy may enhance effort and perseverance in academic activities (Brookhart, 1998; Robbins, Lauver, Le, Davis, & Langley, 2004; Zimmerman, 1995), thus serving as an effective predictor for academic achievement. Another related pair of concepts is based on goal-orientation theory (Ames, 1992); these concepts are mastery and performance orientation. Generally speaking, learners with a mastery goal are more involved in the task and focus on improvement and growth for its own sake, while learners with a performance goal place more emphasis on public recognition and social comparison with others. Although mastery goals are said to be superior to performance goals, the latter are believed to have positive effects as well.

Dörnyei's tripartite framework, as well as other motivational theories mentioned above, addresses the challenge of describing motivation over time (Dörnyei, 2001). A process model of L2 motivation (Dörnyei & Ottó, 1998) was later developed to synthesize various lines of research into a comprehensive framework. In this model, learner motivation is seen as proceeding sequentially through preactional, actional, and postactional phases. Motivational influences under the three phases are different in nature and function differently. In the preactional phase, the learner's choice is determined by various factors shaping their goal setting and intention formation. Many L2 motivation studies focus exclusively on this preactional stage, as exemplified in most studies exploring integration and instrumentation. Learner motivation in the actional phase, however, has scarcely been addressed until now, and is pertinent to the present study on classroom assessment events. According to Dörnyei (2001), once learners move to the actional phase, for example after being enrolled in an L2 course, choice motivation is replaced by 'executive motivation that energizes action while it is being carried out' (p. 85). While learners engage in subtask generation and implementation, they appraise the activity at hand and exercise action control. This appraisal, unlike motivation at the preactional stage, involves the evaluation of a multitude of stimuli, such as learners' perceived quality of the learning experience and self-efficacy for the particular activity. For action control, Dörnyei referred to managing cognitive and non-cognitive resources using knowledge

and strategies for goal fulfillment, particularly emphasizing self-regulatory mechanisms. A discussion of these self-regulatory strategies follows.

In 2006, Tseng, Dörnyei, and Schmitt proposed a new way to scrutinize L2 learning strategies. Instead of following the tradition of L2 learning strategy studies set by 'Strategy Inventory for Language Learning' (SILL, Oxford, 1990), they pointed out the problems of SILL in focusing on the quantity (i.e. number and frequency of strategies) rather than the quality of the strategy used. For example, a learner using only a smaller number of strategies compatible with his or her own learning styles effectively may score lower on the SILL when compared to a less skilled learner trying all strategies randomly. Drawing on work in educational psychology, Tseng et al. (2006) argued that what makes learners strategic is the creative effort they put into trying to improve their own learning. The theoretical construct of self-regulation was then applied in generating a psychometrically based 'Self-regulating Capacity in Vocabulary Learning Scale.' The five areas of identified self-regulation were commitment control, metacognitive control, satiation control, emotion control, and environmental control. Although the scale is designed for the learning of vocabulary, Tseng et al. (2006) view it as a heuristic approach and believe that it could be adapted for use in other language learning domains.

Related empirical studies

In recent years, a few articles in major language measurement journals have started to introduce learning-oriented classroom assessments from other disciplines (e.g. Harlen & Winter, 2004; Leung, 2004; Rea-Dickins & Gardner, 2000). However, empirical research on how classroom assessments may influence learning is relatively underrepresented in the L2 field. The following is a review of empirical studies from other domains of education.

Brookhart and DeVoge (1999), from four classroom assessment events in each of two third grade classrooms, found positive relationships between perceived task characteristics, perceived self-efficacy, amount of invested mental effort, and achievement. The factors influencing these variables were the classroom assessment environment created by the teacher and the number of assessment events embedded in it. The only exception was that, for some assessment events, there was no relationship between perceived self-efficacy and effort. The researchers surmised, based on learner interview data, that some events might have been less challenging such that learners with high self-efficacy did not want to expend effort. They concluded that the relationship between self-efficacy and effort is a complicated one and more research is needed. In addition, they found that students' self-efficacy for particular assessment events is largely based on similar previous experiences. If the results of a previous assessment event were positive, they are accepted as evidence for how well the student may achieve in a similar assessment.

Brookhart and Durkin (2003) investigated three different high school social science classes, each having four different assessment events. They discovered similar evidence. That is, within the same class, with the same teacher and same group of learners, individual learners' perceived task characteristics, perceived self-efficacy, amount of invested mental effort, goal orientation, and strategy use differed from one assessment event to another. Moreover, self-efficacy was stronger with group performance assessments than

with traditional tests. The kind of cooperation opportunities and the associated positive impact on learning provided by group performance assessments was what pencil-and-paper tests failed to offer. The researchers concluded that performance assessments, when compared with tests, are connected with productive goal orientations and learning strategies. In addition, they provide a chance for students of both high and low self-efficacy to achieve. In the same vein, Stefanou and Parkes (2003) researched grade five elementary science courses and compared three types of assessments (i.e. traditional paper-and-pencil tests, laboratory tasks, and performance assessments). They concluded that tests and performance assessments are better at fostering a more task-focused orientation than are laboratory tasks.

Convergent assessment vs. divergent assessment

Terms used in this area of research to describe the nature of assessments are often used interchangeably and inconsistently. Some use 'teacher-based assessment' to emphasize the role of the teacher, such as in the 2009 *TESOL Quarterly* special issue, others use 'classroom assessment' to mean the same thing. 'Performance assessment' is often used to describe distinctive assessment practices that are different from 'tests', but classroom tests can also have a performance element. For the purpose of this study, Torrance and Pryor's (2001) distinction between convergent assessment (CA) and divergent assessment (DA) was found useful. They specify a continuum of possibilities for classroom teachers:

In *convergent assessment*, the important thing is to find out *if* the learner knows, understands or can perform a predetermined concept. It is characterized by detailed planning and generally accomplished by closed or pseudo-open questioning and tasks. In this way, the interaction of the learner with the curriculum is seen from the point of view of the curriculum. The theoretical origins of such an approach would appear at least implicitly to be behaviorist, deriving from mastery-learning models and involving assessment *of* the learner *by* the teacher. *Divergent assessment*, on the other hand, emphasizes the learner's understanding rather than the agenda of the assessor. In this assessment, the important thing is to discover *what* the learner knows, understands, and can do. It is characterized by less detailed planning, where open questions and tasks are of more relevance. The implications of divergent teacher assessment are that a constructivist view of learning is adopted, with an intention to teach in the zone of proximal development (Vygotsky, 1986). As a result, assessment is seen as accomplished jointly by the teacher and the student, and oriented more toward future development rather than measurement of past or current achievement. (Torrance & Pryor, 2001, pp. 616–617)

This pair of definitions successfully removes the fuzziness embedded in other more commonly adopted terms and is applied in this study to represent the two types of classroom assessments observed.

The study

The study adopted a mixed method design. One college EFL (English as a foreign language) teacher and her three groups of students were involved. Each student experienced two classroom assessment events, a CA and a DA. Data on their motivation and strategy under each condition was collected and analyzed. The main research question was, 'Does

student motivation and strategy use differ under CAs as opposed to DAs?’ Furthermore, the researcher examined students’ trait motivation, including integrative motivation, instrumental motivation, requirement motivation, and general self-efficacy in EFL learning and course domain, including speaking and listening, to determine how they interact with motivation and strategy use under different assessment events.

Participants

In a previous research project (Huang, forthcoming) the author interviewed six EFL instructors from a university in northern Taiwan on topics related to classroom assessment practice. One of the six was identified as using both CAs and DAs in her classes. Recruiting all student participants from one single teacher’s classroom eliminated the possible contamination of teacher variation, which was demonstrated to exert more influence than the experiment treatment itself (Yin et al., 2008). This EFL instructor and her three groups of college freshmen, a total of 105 students, were then invited to participate in this study. The instructor held an MA TESOL degree and was a full-time EFL instructor with two years of experience teaching at that university. The university is highly ranked among the island’s universities and students admitted to it are considered to have been academically successful in their previous high school education. Except for only a few who qualify to waive the course, all students are required to take EFL courses in their first two semesters. They can choose two out of three skill concentration alternatives, namely reading, listening, or speaking. Once a skill is chosen, students are randomly assigned to different classes of approximately 30 to 50 students each. Student participants in this study were those enrolled in one listening and two speaking classes taught by the instructor.

Description of the assessment events and data collection procedures

Apart from the students involved, the two speaking classes in this study were identical in all respects – course design, syllabus, and procedures. In both speaking sections, students took an oral exam (i.e. the CA) in which they came on the same day in groups of three to answer questions and interact with each other under the instructor’s supervision. The DA was a group presentation on a specific topic for which student groups prepared beforehand according to the instructions given. They then gave an oral presentation in front of the class. This type of group presentation has been commonly used in EFL courses, as already shown in a number of survey studies (Cheng, Rogers, & Hu, 2004; Ricardo-Osorio, 2008). Presentations were scheduled throughout the second half of the semester.

In the listening section, students took a listening test (i.e. the CA) in a format similar to TOEIC listening tests, except that it was shorter and the materials were related to what students had studied in the course. Students in this course also had a group presentation (the DA) for which they preselected a video clip they had searched for on the Internet that had no subtitles. They then collaboratively produced subtitles and submitted the work on the course e-platform for the instructor to view and check before class time. For the presentation, they played the video to the class for two minutes and subsequently introduced the video and learnable English expressions verbally. As with the speaking sections, the CA was done in one class meeting with all students and the DA was scheduled throughout the second half of the semester.

Each participating student filled out three questionnaires, all with statements anchored on a Likert-type scale ranging from 1 (unimportant, bad, disagree) to 6 (important, good, agree). The first questionnaire tapped into students' trait motivation and was administered at the beginning of the term. The other two questionnaires focused on event-specific motivation and strategy, one for the CA and the other for the DA. On the day after students had completed the event, either a CA or DA, they were asked to fill out the relevant questionnaire. All students received a piece of stationery worth US \$3 from the researcher as a token of gratitude. Details of the questionnaire items are described in the following section.

Instruments

The trait questionnaire (as shown in Appendix A) consisted of four subscales, integrative motivation (INT), instrumental motivation (INS), requirement motivation (REQ), and self-efficacy (SEF), each of which contained four items. Most of the motivational orientation items were adapted from Chen, Warden, and Chang (2005) because their items had been used on a similar learner population. The three motivational orientations have been studied extensively in the language learning motivation literature and self-efficacy has been demonstrated to be an important factor in determining student effort and strategy use. Therefore, these variables were investigated as covariate candidates in the comparison of effects between CA and DA.

The event-specific questionnaires (as shown in Appendix B) were parallel under CA and DA, and each contained two parts, motivation and strategy. The motivation part included five subscales – perceived task characteristics (PTC, 5 items), perceived self-efficacy (PSE, 5 items), amount of invested mental effort (AIME, 3 items), mastery goal orientation (MGO, 2 items), and performance goal orientation (PGO, 3 items); items were mainly adapted from Brookhart, Walsh, and Zientarski (2006). For the learning strategy part, the researcher applied Tseng et al.'s (2006) five categories of self-regulated learning strategies. These are commitment control (COM), metacognitive control (MET), emotional control (EMO), satiation control (SAT), and environment (ENV) control. Each strategy subscale consisted of four items.

Results

Reliability and validity of scales

The Cronbach α for the entire questionnaire was .865. For trait variables, the α values were .725, .826, .807, .839 for INT, INS, REQ and SEF respectively. For each event-specific variable, reliability coefficients were obtained for both CA and DA, and the mean alphas were .803 for PTC, .697 (PSE), .933 (AIME), .767 (MGO), .827 (PGO), .740 (COM), .824 (MET), .422 (EMO), .638 (SAT), and .770 (ENV). Scale reliability was satisfactory except for the subscales of EMO and SAT; therefore, they were removed from follow-up analysis. PSE was somewhat low, but it was retained because of the significance of its contribution often found in previous studies.

Validity of trait variables was examined by performing a factor analysis for all 16 items involved. Results indicated that the items clustered nicely into four distinct factors as expected, as shown in Table 1.

Table 1. Rotated factor matrix for trait questionnaire items

	Item	1	2	3	4
Instrumental motivation	1	−0.0014	0.8562	0.1112	0.1207
	2	−0.0295	0.8674	0.1362	0.0436
	3	−0.1475	0.7296	0.1006	0.0993
	4	−0.0151	0.7200	0.2620	0.1589
Requirement motivation	5	0.0190	0.2548	0.7091	0.1005
	6	−0.0501	0.1228	0.8825	−0.0464
	7	−0.0055	0.0316	0.8518	0.1160
	8	0.0385	0.3490	0.5960	0.2156
Integrative motivation	9	−0.0403	0.1121	−0.0233	0.7433
	10	0.0252	0.1498	0.0613	0.8090
	11	0.0313	0.4294	0.2383	0.5517
	12	0.0934	−0.0062	0.1437	0.7159
Self-efficacy	13	0.9063	−0.0646	0.0512	−0.0165
	14	0.8930	−0.0473	0.0252	−0.0053
	15	0.9249	−0.0477	−0.0018	0.0537
	16	0.7682	−0.0132	−0.0722	0.0669

Extraction: Principal component
Factor rotation: Varimax rotation with Kaiser Normalization

Validity of event-specific subscales was examined by seeking convergent and discriminant validity evidence using a multi-trait multi-method matrix (MTMM). The result is shown in Table 2. As can be seen in the table, each figure in the cells is a correlation coefficient of the subscales in its respective row and column, except that those in the diagonal are the subscales' Cronbach α . Figures in the shaded area (i.e. the scales' convergent validity) are smaller than their respective α . In addition, they are larger than the corresponding discriminant validity figures (i.e. all other figures in the same row and column). For example, the convergent validity figure of PTC, 0.399, is smaller than the Cronbach α of CA (0.884) as well as that of DA (0.722). However, 0.399 is larger than all PTC's corresponding discriminant validity figures in its row (0.105, 0.292, 0.245, 0.090, 0.086, 0.081, 0.284) and its column (0.285, 0.262, 0.369, 0.131, 0.201, 0.205, 0.243). This validity check permitted us to proceed with the analysis.

Descriptive statistics

The descriptive statistics for trait variables are presented in Table 3 and those for event-specific variables in Table 4. Because an initial examination indicated that results from speaking and listening sections showed different patterns, the results from these two sections were presented separately. All figures were calculated as the average of all items using the same scale ranging from 1 to 6. From Table 3, a general student motivation profile is apparent. These participants are relatively highly motivated in L2 learning, with integrative motivation at 5.25, instrumental motivation at 5.16, and requirement motivation at 4.87. The more interesting part is their overall self-efficacy. Compared to the three motivation scales, their self-efficacy is much lower with a much larger standard deviation.

Table 2. MTMM for all subscales

Items	CA	DA														
	PTC	PSE	AIME	MGO	PGO	COM	MET	ENV	PTC	PSE	AIME	MGO	PGO	COM	MET	ENV
CA																
PTC	0.722															
PSE	0.404	0.637														
AIME	0.411	0.115	0.926													
MGO	0.513	0.204	0.633	0.753												
PGO	0.092	0.224	-0.050	0.012	0.766											
COM	0.186	0.287	0.196	0.220	0.085	0.693										
MET	0.150	0.296	0.182	0.158	0.009	0.684	0.773									
ENV	0.402	0.185	0.320	0.453	0.019	0.481	0.371	0.688								
DA																
PTC	0.399	0.105	0.292	0.245	0.090	0.086	0.081	0.284	0.884							
PSE	0.285	0.442	0.250	0.210	0.202	0.194	0.184	0.170	0.532	0.757						
AIME	0.262	0.047	0.607	0.444	0.020	0.162	0.160	0.342	0.607	0.564	0.939					
MGO	0.369	-0.007	0.431	0.528	-0.054	0.101	0.034	0.337	0.592	0.365	0.665	0.781				
PGO	0.131	0.192	0.088	0.170	0.600	-0.056	-0.130	-0.037	0.281	0.483	0.325	0.319	0.888			
COM	0.201	0.104	0.254	0.315	-0.015	0.480	0.392	0.470	0.485	0.487	0.571	0.533	0.256	0.786		
MET	0.205	0.124	0.013	0.129	0.010	0.410	0.521	0.352	0.461	0.437	0.415	0.389	0.167	0.753	0.875	
ENV	0.243	-0.067	0.306	0.280	0.054	0.398	0.232	0.713	0.522	0.256	0.510	0.480	0.110	0.637	0.484	0.852

Note: CA = convergent assessment, DA = divergent assessment, PTC = perceived task characteristics, PSE = perceived self-efficacy, AIME = amount of invested mental effort, MGO = mastery goal orientation, PGO = performance goal orientation, COM = commitment control, MET = metacognitive control, ENV = environmental control

Table 3. Descriptive statistics for trait variables (in a scale from 1 to 6)

Scale (no. of items)	<i>n</i>	<i>range</i>	<i>Min.</i>	<i>Max.</i>	<i>Mean</i>	<i>SD</i>
INT (4)	105	2.75	3.25	6.00	5.25	.62
Speaking	61	2.75	3.25	6.00	5.32	.62
Listening	44	2.25	3.75	6.00	5.16	.62
INS (4)	105	2.75	3.25	6.00	5.16	.69
Speaking	60	2.75	3.25	6.00	5.25	.69
Listening	45	2.25	3.75	6.00	5.04	.69
REQ (4)	102	3.00	3.00	6.00	4.87	.69
Speaking	58	3.00	3.00	6.00	4.91	.70
Listening	44	2.75	3.25	6.00	4.82	.69
SEF (4)	104	4.75	1.00	5.75	3.73	.81
Speaking	61	4.75	1.00	5.75	3.54	.91
Listening	43	3.00	2.25	5.25	4.00	.53

Note: INT = integrative motivation, INS = instrumental motivation, REQ = requirement motivation, SEF = overall EFL self-efficacy

Among the four trait variables, a *t*-test showed that the two groups of students in the speaking and listening sections differed in their overall self-efficacy, but not in the three motivational traits. The listening group’s self-efficacy was higher ($t = 2.984, p = .004$).

Table 4 shows the means and standard deviations of each scale under different assessment events in the two types of courses. The mean differences and paired-sample *t*-tests between two events were also calculated and presented in the same table. The general pattern observed in this table is that, before any trait variable is controlled for, the event-specific variables generally weighed favorably for one assessment event over the other. That is, they were uniformly higher or lower under CAs or DAs. Specifically, the measures were higher in DAs for speaking and in CAs for listening, with only a few exceptions.

Correlation between trait and event-specific variables

In order to find out which trait variables had a stronger impact on event-specific motivation and strategy use, and should consequently be controlled for as covariates, the correlation between trait and event-specific variables was scrutinized. The results are shown in Table 5. The left column represents the difference of event-specific subscales of each participant for the CA and the DA situations. The top row represents the four trait variables. The figures in the table are correlation coefficients between these two. Results suggest that none of the three motivational orientations are related to the difference between CA and DA. However, overall self-efficacy is significantly related to the difference between the two assessment events for five out of the eight subscales. In particular, the relationship is a negative one. That is to say, for students with lower overall self-efficacy, their motivation and strategy (PTC, PSE, MGO, COM, and ENV) were higher with DAs than with CAs. But for those with higher overall self-efficacy, their motivation and strategy use were higher in the CA situation than in DA. Based on the results in Table 5, SEF was chosen as a covariate in the comparison of assessment events.

Table 4. Unadjusted descriptive statistics for motivation and strategy scales under convergent and divergent assessments in the speaking and listening sections

Scale (no. of items)	Speaking						Mean Diff.	df	Paired-sample t
	CA			DA					
	M	SD	n	M	SD	n			
Motivation									
PTC (5)	4.14	0.68	61	4.29	0.74	61	-0.14	60	1.548
PSE (5)	3.48	0.77	60	3.81	0.61	60	-0.33	59	-3.694**
AIME (3)	4.30	0.94	60	4.58	0.73	60	-0.28	59	-2.951**
MGO (2)	4.49	0.81	60	4.64	0.70	60	-0.15	59	-1.539
PGO (3)	3.36	0.86	60	3.66	1.04	60	-0.31	59	-2.659*
Strategy									
COM (4)	4.19	0.67	56	4.38	0.60	56	-0.18	55	-2.971**
MET (4)	3.86	0.73	57	4.19	0.77	57	-0.33	56	-4.816**
SAT (4)	3.85	0.69	57	4.06	0.62	57	-0.21	56	-3.125**
EMO (4)	3.97	0.64	57	3.97	0.63	57	0.00	56	.000
ENV (4)	4.48	0.63	56	4.46	0.75	56	0.02	55	.343
	Listening						Mean Diff.	df	Paired-Sample t
	CA			DA					
	M	SD	n	M	SD	n			
Motivation									
PTC (5)	4.04	0.58	42	3.32	0.59	42	0.72	41	6.730**
PSE (5)	3.76	0.72	44	3.32	0.67	44	0.45	43	4.673**
AIME (3)	3.98	0.88	44	3.60	0.77	44	0.39	43	3.400**
MGO (2)	3.67	0.74	44	4.16	0.87	44	-0.49	43	-4.110**
PGO (3)	3.41	0.87	44	3.01	0.93	44	0.40	43	4.179**
Strategy									
COM (4)	4.23	0.69	41	3.75	0.44	41	0.48	40	4.296**
MET (4)	4.13	0.74	42	3.55	0.67	42	0.58	41	5.383**
SAT (4)	3.95	0.62	42	3.71	0.58	42	0.24	41	2.451**
EMO (4)	4.13	0.60	42	3.87	0.46	42	0.26	41	2.477**
ENV (4)	4.39	0.55	42	3.81	0.38	42	0.58	41	8.311**

* $p < .05$; ** $p < .01$

Notes: CA = convergent assessment, DA = divergent assessment, PTC = perceived task characteristics, PSE = perceived self-efficacy, AIME = amount of invested mental effort, MGO = mastery goal orientation, PGO = performance goal orientation, COM = commitment control, MET = metacognitive control, SAT = satiation control, EMO = emotion control, ENV = environment control

The comparison

A one-way multivariate analysis of covariance (MANCOVA) was conducted to determine if a significant difference existed between the two types of classroom assessment events (CAs and DAs) in two different courses. The independent variable was course domain at two levels (listening vs. speaking). The covariate was participants' overall

Table 5. Correlation between trait variables and the difference in motivation and strategy between two types of assessment

Scale difference ^a	INT	INS	REQ	SEF
PTCd _a –PTCc _a	.125	.130	.060	–.244*
PSEd _a –PSEc _a	.059	.167	.054	–.410**
AIMEd _a –AIMEc _a	.122	–.003	–.111	–.179
MGOd _a –MGOc _a	–.146	.018	–.016	–.213*
PGOd _a –PGOc _a	–.067	.095	–.021	–.039
COMd _a –COMc _a	–.031	.094	.109	–.265**
METd _a –METc _a	.003	–.031	–.009	–.155
ENVd _a –ENVc _a	–.107	.162	–.011	–.323**

^a Calculated by subtracting each participant's subscale scores (PTC, PSE, AIME, etc.) under the test situation from those under the performance assessment situation. For example, PTCd_a–PTCc_a = scores of perceived task characteristics in divergent assessment minus scores of perceived task characteristics in convergent assessment.

* $p < .05$, ** $p < .01$

self-efficacy measured at the beginning of the term. The dependent variables, a within-subject repeated measure, were the difference in each participant's motivation and strategy use between CA and DA (e.g. PTC under DA minus PTC under CA). Results are summarized in Table 6. Analyses of variances (ANOVA) on each dependent variable were conducted as follow-up tests to the MANCOVA. Using the Bonferroni method, each ANOVA was tested at the .00625 (.05 divided by 8 dependent variables) significance level. All eight pairs of comparison yielded significant p values smaller than .000.

Conclusion and discussion

Three points are observed in the results. First, students perceived and reacted to CA and DA differently to a significant level in both speaking and listening courses. Second, students' overall self-efficacy, but not any of the three trait motivations (integrative, instrumental, and requirement), had an impact on how they reacted to CA versus DA. In particular, less efficacious students tended to demonstrate higher motivation and strategy use under DA and more efficacious students did so under CA. Third, after having controlled the factor of overall self-efficacy, learner motivation and strategy use showed

Table 6. MANCOVA Summary

Effect	Wilks' λ	F	Hypothesis df	Error df	Sig.	Partial η^2
Intercept	.853	1.763	8	82	.096	.147
Covariate (SEF)	.820	2.251	8	82	.032*	.180
Domain (L vs. S)	.447	12.682	8	82	.000**	.553

* $p < .05$; ** $p < .01$

SEF = self-efficacy; L = Listening; S = Speaking

different patterns under different course domains. In the speaking sections, student motivation and strategy use were higher under DA. But such a pattern was reversed in the listening course. The first point supports Brookhart and DeVoge (1999) in that different classroom assessment events bring about different learner motivation and strategy use, and that these dependent variables generally move in the same direction. That is, when learners perceived an assessment event as more valuable and themselves more efficacious, they were more motivated and expended more effort. However, the second point contradicts previous findings and the third casts some doubt on our previous understanding. We now turn to a discussion of these findings.

Brookhart and Durkin (2003) asserted that performance assessments (PAs) are superior to tests in terms of productive goal orientation and learning strategies, because they provide opportunities for both high- and low-efficacious students to achieve. The second conclusion in this study for learners with lower self-efficacy does support Brookhart and Durkin (2003), but the same does not hold for high-efficacious students. On the contrary, and in spite of our intuitions, a convergent form of assessment, or what we often refer to as tests, works better for high-efficacious students, provided that motivation and strategy use is our focus of concern.

What was observed very likely has to do with learners' cultural background and learning history. Unlike students in previous studies in Western contexts, the participants here were college freshmen in Taiwan who had experienced highly competitive and test-oriented education in the previous six years of secondary school, surviving nationwide entrance examinations both for senior high and college. Despite national endeavor in redesigning entrance exams to alleviate pressure, students' schedules are still dominated by daily quizzes in their own classes, weekly and monthly exams at the school level, and mock review exams imitating the format of national exams and covering various ranges in their school districts. Almost all exam results are numerical and normative, showing each single student's standing in relation to his or her peers. The concept of percentile ranking is not limited to testing professionals; it has been a language for communication among teachers, students, and parents. For example, most ninth graders (those in the year prior to entrance exams for senior high, when students anxiously count down to the final exam day) and their parents are able to report the student's 'PR values' (as from the percentile figure, e.g. 97 or 68) from the most recent exam and use this information to predict which school the learner will be admitted to. Those who always outperform others are encouraged and take pride in their achievement in a normative sense. The aforementioned societal phenomenon may provide us some clues as to why more efficacious students had higher motivation and better self-regulatory strategies in the convergent individual tests than in the more divergent group performance.

The third point above also casts doubt on a previous research finding (Brookhart & Durkin, 2003; Stefanous & Parkes, 2003) that the more divergent form of assessment, or what has been commonly referred to as performance, is better than convergent or test-like assessments in inducing motivation and strategy use. By perusing the present results, one can see that the conclusion was supported in the speaking course but questioned in the listening course. In addition to the issue of educational culture explicated above, it is therefore necessary to revisit and scrutinize the specific assessment events. For the speaking sections, both the DA, a group presentation on a topic of the student's choice,

and the oral test, in which students answered the questions selected by the teacher from a question bank for which they had practiced, were both quite typical. The higher order thinking and depth of engagement required by the open-ended DA task were probably what caused the higher motivation and strategy use compared to the CA task, which involved broader coverage of materials selected by the teacher rather than by the student.

Results from the listening section were more surprising. Why was a CA better received by students than a DA? Why were the listening section results different from what was observed in speaking and in other situations reported by previous studies? First of all, a listening DA may be less common. For this receptive and unobservable skill, some common authentic tasks may involve note-taking, but it is rare for students to have to give a presentation on what they have listened to. However, what this particular instructor designed in the listening DA was innovative and had clear face validity. Under her instruction, students had to search for video clips and write subtitles for them. After an expected effort of authentic and intensive listening, they would then be able to show the results and give the presentations.

One might ask why this DA produced lower motivation and strategy than a CA. First, according to Brookhart and DeVoge (1999), 'students' self-efficacy judgments about their abilities to do particular classroom assessments were based on previous experiences with similar kinds of classroom assessments' (p. 422). Because of the lack of similar prior experience on this particular DA, especially when it was compared with a more traditional listening comprehension test, the CA, students demonstrated lower perceived task characteristics, lower self-efficacy, lower amount of invested mental effort, and lower performance goals. The higher mastery goal (as shown in Table 4) may be a sign of the potential of such a DA once the novelty factor is eliminated, however. As to strategy, though, all measures were uniformly lower in the DA situation. Learners did not use more self-regulated learning strategies under the less familiar assessment situation. Although past studies seem to suggest that the absence of previous evidence of success may induce more student effort, that proposition was made with some reservation (Brookhart & DeVoge, 1999). The relationship between motivation and strategy use under various assessment situations appears to be a complicated one.

The above findings supplement those of previous studies. The subject matter (EFL), the students' age and educational level (college freshmen), and the social milieu (the competitive test-oriented and norm-referenced educational climate) are the new elements this study has brought to bear. It also has to be noted that the four particular assessment events observed were not designed for this study, but sampled from real classrooms. There was no intervention as far as how the participating EFL teacher designed her instruction and assessments. These assessment events served as a showcase of her beliefs, past training, and personal innovations.

Limitations, future research directions, and pedagogical implications

The following limitations have to be considered when conclusions are drawn; they also indicate future possible research directions. The first one has to do with the variation between participants. The course domain, speaking versus listening, was a between-subject measure because the participants were only involved in one of them. It was found

that the two groups differed in self-efficacy, which was in turn controlled for by the MANCOVA analysis. However, it was not clear if these two groups differed significantly in other ways and whether that difference may be a factor in the patterns of motivation and strategy. Future studies should consider a design in which between-subject variations may be removed.

Second, because the data collected was based solely on questionnaires, in-depth qualitative data, such as interviews with learners on whether they perceived these assessments as high or low stakes, were lacking. However, these perceptual factors may have played a critical role. Future studies trying a more qualitative approach may help us gain better insight. Moreover, the assessment events, especially the divergent ones, were at the same time instructional activities. Learners may take the same assignment as either a learning opportunity and focus on the learning itself, or feel some pressure to gain a high score and try things other than what the teacher originally intended, as shown in Spence-Brown's (2001) study of task authenticity. Would they react more positively or negatively if grades were not assigned? Would the grade given matter, and to what extent? These are the elements that should be incorporated into the design of future studies to help us understand the issue better.

Third, is skill integration an issue in our understanding of classroom assessments? Both the listening DA and the speaking CA were integrated activities involving more than one skill. In a strict testing tradition such practice may endanger test validity. However, the integrative approach is accepted and not new in language testing, as exemplified in large-scale proficiency tests such as iBT TOEFL. The more relevant question here is, 'What is the impact of skill integration, as compared to single skill assessment events, on learner motivation and strategy use?' Does the integration make a difference, and, if the answer is positive, to what extent? These questions point to another future research direction.

Finally, a discussion on pedagogical implications follows. Divergent assessments, as specified by Torrance and Pryor (2001), are intended to teach within the learners' zone of proximal development (Vygotsky, 1986), and require the collaboration of the learner with teachers and peers. They help many learners experience success that is usually unattainable in the more traditional convergent, test-like, assessments. Now, with the dissemination of the *assessment for learning* movement in more fields of education, teachers are gradually liberated from large-scale standardized testing and top-down assessment policies, and at the same time given more responsibilities for teaching and assessing students. How to design assessment events that can facilitate learning, and are meaningful and motivating, becomes an issue more prominent than ever before. The ultimate question is, 'If we want to take full advantage of assessment opportunities for our teaching and student learning, what are the new possibilities that have been overlooked in the past, and what potential is associated with these opportunities?' We want to make sure that these opportunities are beneficial in long-term learning and proficiency gains. But before achievement can be reliably measured by us and felt by students, they must be engaged in the assessment events and be motivated to put forth more effort; then we can reasonably expect proficiency gains to come naturally from motivation and strategy use.

Acknowledgements

This paper is a result of a research project sponsored by the National Science Council in Taiwan, Grant #NSC96-2411-H-004-047. My special thanks go to the anonymous reviewers for their insightful comments on earlier drafts of this article.

References

- 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, 359–382.
- Ames, C. (1992). Classrooms, goals, structures and student motivation. *Journal of Educational Psychology*, 84, 267–271.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, 84, 191–215.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning – Putting it into practice*. Berkshire, UK: Open University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, September, 9–21.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan*, October, 139–148.
- Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10(2), 161–180.
- Brookhart, S. M. (1998). Determinants of student effort on schoolwork and school-based achievement. *The Journal of Educational Research*, 91, 201–208.
- Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, 12(4), 409–425.
- Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27–54.
- Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education*, 19, 151–184.
- Butler, Y. G. (2009). How do teachers observe and evaluate elementary school students' foreign language performance? A case study from South Korea. *TESOL Quarterly*, 43(3), 417–444.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, September, 40–43.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21, 360–389.
- Chen, J. F., Warden, C. A., & Chang, H.-T. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly*, 39, 609–633.
- Connor-Greene, P. A. (2000). Assessing and promoting student learning: Blurring the line between teaching and testing. *Teaching of Psychology*, 27(2), 84–88.

- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90–100.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415.
- Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal*, 78, 273–284.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Pearson Education Limited.
- Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics*, 4, 43–69.
- Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, 40(2), 117–128.
- Fox, J. (2004). Biasing for the best in language testing and learning: An interview with Merrill Swain. *Language Assessment Quarterly*, 1(4), 235–251.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Hamp-Lyons, L. (2009). Principles for large-scale classroom-based teacher assessment of English learners' language: An initial framework from school-based assessment in Hong Kong. *TESOL Quarterly*, 43(3), 524–529.
- Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing*, 21, 390–408.
- Huang, S.-C. (forthcoming). Pushing learners to work through tests and marks: Motivating or demotivating? – A case in a Taiwanese university. *Language Assessment Quarterly*.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 49–74.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly*, 1(1), 19–41.
- Leung, C. (2007). Dynamic assessment: Assessment *for* and *as* teaching? *Language Assessment Quarterly*, 4(3), 257–278.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333–349.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667–686.
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, 91(3), 323–340.
- Poehner, M., & Lantolf, J. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9, 233–265.
- Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17, 215–243.
- Ricardo-Orsorio, J. G. (2008). A study of foreign language learning outcomes assessment in U.S. undergraduate education. *Foreign Language Annals*, 41(4), 590–610.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288.

- Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, 18(4), 463–481.
- Stefanou, C., & Parkes, J. (2003). Effects of classroom assessment on student motivation in fifth-grade science. *The Journal of Educational Research*, 96(3), 152–162.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163–184.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615–631.
- Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78–102.
- Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: Harvard University Press.
- Warden, C. A., & Lin, H. J. (2000). Existence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, 33, 535–547.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M. et al. (2008). On the impact of formative assessment on student motivation, achievement and conceptual change. *Applied Measurement in Education*, 21, 335–359.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.) *Self-efficacy in changing societies* (pp. 202–231). Cambridge: Cambridge University Press.

Appendix A: Items from the trait questionnaire

How important is English to you in pursuing the following circumstances?

Instrumental motivation (INT)

1. Obtain a higher paying job
2. Obtain a raise
3. Change jobs more easily
4. Have greater job security

Requirement motivation (REQ)

5. Pass an examination for further study
6. Pass a required class
7. Pass an elective class
8. Pass an exam for a job position

Integrative motivation (INT)

9. Travel overseas
10. Make social contacts

11. Gain social prestige
12. Understand foreign movies, books, and magazines

Self-efficacy (SEF)

13. How good is your English ability?
14. ...compared to peers in the same class?
15. ...compared to peers in the same university?
16. ...compared to young people of the same age?

Appendix B: Items from the four situation-specific questionnaires (X being one of the four assessment events)

Motivation

Perceived task characteristics (PTC)

1. I am very interested in X.
2. It is important for me to learn [about] X.
3. It is useful for me to learn [about] X.
4. I like X.
5. I think I will be able to use what I learned from X in other work.

Perceived self-efficacy (PSE)

6. X was difficult for me.
7. I think I will do well on X compared to other work I do in this class.
8. I think I will do well on X compared to other students in this class.
9. I believe I will receive an excellent grade for X.
10. I'm confident I understood the most complex part of the work involved in X.

Amount of invested mental effort (AIME)

11. I tried very hard when I did X.
12. I concentrated a lot while I did X.
13. I put a lot of effort into preparing for X.

Mastery goal orientation (MGO)

14. For X, I wanted to learn as much as possible.
15. I wish we had more time to spend on X.

Performance goal orientation (PGO)

16. I wanted others to think I was smart at X.
17. It was important to me to do better than the other students on X.
18. It was important to me that the teacher thought I did a good job on X.

Self-regulated strategies (X being preparing for one of the four assessment events)

Commitment control (COM)

1. When X, I have special techniques to achieve my learning goals.
2. When X, I believe I can achieve my goals more quickly than expected.
3. When X, I persist until I reach the goals that I made for myself.
4. I believe I can overcome all the difficulties related to achieving my X goals.

Metacognitive control (MET)

5. When X, I have special techniques to keep my concentration focused.
6. When X, I think my methods of controlling my concentration are effective.
7. When it comes to X, I have my special techniques to prevent procrastination.
8. When it comes to X, I think my methods of controlling procrastination are effective.

Satiation control (SAT)

9. Once the novelty of X is gone, I easily become impatient with it (reversely coded).
10. During the process of X, I feel satisfied with the ways I eliminate boredom.
11. During the process of X, I am confident that I can overcome any sense of boredom.
12. When feeling bored with X, I know how to regulate my mood in order to invigorate the learning process.

Emotion control (EMO)

13. When I feel stressed about X, I know how to reduce this stress.
14. I feel satisfied with the methods I use to reduce the stress of learning in X.
15. When I feel stressed about X, I simply want to give up (reversely coded).
16. When I feel stressed about X, I cope with this problem immediately.

Environmental control (ENV)

17. When I am X and the learning environment becomes unsuitable, I try to sort through the problem.
18. When X, I know how to arrange the environment to make learning more efficient.
19. When X, I am aware that the learning environment matters.
20. When X, I look for a good learning environment.