

# Illustrating Students' Perceptions of English Language Assessment: Voices from China

RELC Journal  
0(0) 1–22  
© The Author(s) 2013  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0033688213500595  
rel.sagepub.com  
The SAGE logo consists of a stylized 'S' inside a circle, followed by the word 'SAGE' in a bold, sans-serif font.

**Yangyu Xiao**

The University of Hong Kong, Hong Kong

**David Robert Carless**

The University of Hong Kong, Hong Kong

## Abstract

Whilst there are studies exploring students' perspectives of assessment, relatively little is known about how Chinese high school students' respond to assessment. Students' perspectives are particularly important because the power of assessment directly impacts on their motivation and attitudes towards schooling. This article explores students' views of English language assessment in a Chinese high school in Hubei province through 'draw-a-picture technique' and subsequent interviews. The main findings show that assessment could invoke both positive and negative feelings; and that tests carried both summative and formative messages. The paper contributes to the understanding of how assessment impacts on students' lives and discusses the use of draw-a-picture technique as a means of documenting students' perceptions.

## Keywords

Assessment, student perceptions, drawings

## Introduction

It has been recognized for a number of decades that assessment has a profound impact on how students experience schooling and how they perceive themselves as language learners and as individuals (e.g. Shohamy, 2007). The experiences and perceptions of students are also impacted by the educational and social context in which they are studying.

---

### Corresponding author:

Yangyu Xiao, 109 Hui Oi Chow Science Building, Faculty of Education, The University of Hong Kong, Pokfulam, 999077, Hong Kong.  
Email: shirleyxiaoyy@gmail.com

Accordingly, there is a need to situate understandings of assessment practices within wider socio-cultural settings (Xu and Liu, 2009), which in China includes particular aspects of context, culture and history. Competitive examinations, for example, date back two thousand years to the Han dynasty and carry residual influences on contemporary assessment practices (Carless, 2011).

Students in China are particularly frequently tested (Cheng, 2008). They are involved in internal tests and examinations; competitions with other local or regional schools; and most fundamentally, large-scale high-stakes examinations which deeply impact on their future life chances. There are, however, few major studies which document how Chinese high school students respond to testing and assessment. How do different students react to preparing for and taking tests? To what extent do tests lead to student satisfaction and achievement or pressure and discouragement? How and why do some students derive some formative orientation from assessment, whereas many seem to perceive it as mainly carrying a summative function?

The aim of this exploratory study is to analyse students' perceptions of their experiences of English language assessment in the Chinese high school context. Participants were students (aged 16 to 17) in a school in Hubei province in central China who were working towards the National Matriculation English Test (NMET) which is the main channel for university entrance. Students expressed their perceptions of assessment through drawings and related in-depth interviews. The paper contributes to ELT research through two aspects. First, it enriches understandings of high school students' perspectives of English language assessment in China and extends the analysis of how student perceptions are influenced by aspects of the educational culture to which they belong. Second, a methodological contribution is to build on recent interest in the use of student drawings in assessment research (Brown and Wang, 2011; Carless and Lam, 2012; Harris, et al., 2009) and share this technique with the wider ELT community.

## **The Power of Testing and Its Affective Impact**

The framework for the paper encompasses three related strands: the power of tests; students' affective responses to assessment; and the interplay between summative and formative assessment. First, we define key terms used in the paper. Assessment is a super-ordinate term for all forms of assessment, whereas testing denotes one particular form of assessment (Leung and Lewkowicz, 2006). Summative assessment mainly has a function of summing up student achievement at a certain point in time, whereas formative assessment is focused upon the development of student learning and can involve, for example, support from teachers, peers or self-reflection (Carless, 2011).

### *The Power of Tests*

As the results of high-stakes tests often facilitate or reduce future study opportunities for individuals, they exert significant impact on test-takers. The power and status of tests encourages teaching to the test and from a student perspective, test preparation can evoke feelings of fear, unfairness, bias, pressure and suspicion (Shohamy, 1998). Data from Israeli high school students, for example, indicate that they doubt whether test scores provide a full

picture of their language abilities (Shohamy, et al., 1996). Matriculation tests in China have also been found to exert strong washback on students' learning, for example, their way of practicing reading, and their way of coping with test passages (Xiao, et al., 2011).

Through its power, assessment also impacts on students' identities as learners. The experiences of seven students taking their matriculation test in Finland were analysed on the basis of oral diaries gathered with portable recording devices over a four month period (Huhta, et al., 2006). The study found that students' perceptions were complex and multi-faceted in that they expressed varied emotions during a single episode and over time. Students revealed different roles within a series of repertoires: hard-working or lazy, skilled or unskilled, cool or nervous, lucky or unlucky.

The power of examinations is recognized in Chinese schools from an early age and test-preparation activities often dominate classroom processes, especially towards the end of senior high school (Jin and Cortazzi, 2006). Li and Baldauf (2011) observe that in ELT in China test scores are treated by most teachers as the main or only purpose for student learning. For most Chinese students, tertiary education in China is regarded as the only pathway to a successful career and a more promising life (Cheng and Qi, 2006). With limited places available, the large number of candidates accentuates the powerful consequences and the competitive nature of examinations (Qi, 2010).

### *Student Affective Response*

The second strand of our framework concerns student affective responses to testing. Given the power of testing and its impact on students' future prospects, it is not surprising that it has a deep influence on students' feelings and emotions. Pressure on students generally intensifies in high school as high-stakes terminal school examinations approach. Anxiety, for example, is reported by 96% of a sample of Israeli grade 9 to 12 students preparing for an oral English test as part of their matriculation examinations (Shohamy, et al., 1996). Assessment can also provide motivation and encouragement when results are good. Satisfaction often follows from positive experiences, such as praise or good marks (Dörnyei and Ushioda, 2011). Successful performance can also lead to feelings of pride and a sense of achievement (Oxford and Shearin, 1994).

Affective responses to assessment have also been explored in Chinese contexts across students of different age groups. Data from 115 lower elementary school students from nine schools in Hong Kong showed that the majority of students perceived assessment negatively, whilst a significant minority expressed the view that assessment could make them happy when results were good (Carless and Lam, 2012). A recent questionnaire study completed by 89 mainland Chinese high school learners (Xiao, et al., 2011) illustrated students adopting various strategies to tackle test anxiety, such as slowing down and taking deep breaths; students also reported feeling more confident in facing examinations after doing more test preparation.

### *Interplay between Summative and Formative Assessment*

The third strand of our framework concerns the interplay between summative and formative assessment. The enduring influence of summative assessment in official and

professional discourse tends to undermine formative assessment (Leung and Rea-Dickins, 2007). Even when there are efforts to promote formative assessment at a governmental level, the prevailing dominance of summative assessment is a barrier to changes in stakeholders' mindsets (e.g. Butler, 2009; Deng and Carless, 2010). For example, despite the new syllabus in China encouraging the promotion of formative assessment (Wang, 2007), the ubiquitous presence of summative assessment exerts its impact at all levels of schooling.

Our position, congruent with some recent literature (Carless, 2011; Davison and Leung, 2009), is that although summative assessment is often a barrier to formative assessment, there may also be scope for positive interplay between the two. This mainly involves the possibility of some formative potential from processes that might be interpreted by students as largely summative (Davison, 2007). Using case study data from ELT classes in elementary schools in Hong Kong, Carless (2011) illustrates how teachers develop some formative potential from testing processes, for example, asking students to carry out self-reflection on test performance for future improvement.

### *Summary of Framework*

To sum up, our framework suggests that tests exert power over individuals in terms of their life chances and how they are viewed by others. Results from summative assessments induce a variety of affective responses which could include, for example, satisfaction, sense of achievement, pressure or anxiety. The dominance of summative assessment tends to undermine formative assessment, although there are some possibilities for productive interplay between the two. Our study seeks to document how a group of Chinese high school students responds to assessment.

### **Context of the Study**

Students' perceptions of assessment cannot be adequately understood without taking into account the educational context to which they belong. Accordingly, this section establishes the context for the study by introducing selected aspects of education in China and the local Hubei setting.

School education in China encompasses nine-years of compulsory education (usually six-years of elementary and three-years of junior secondary) which all students are required to complete and three-years of high school education where the enrollment is based on the performance in the municipal-level high-school entrance examinations. ELT in high schools is focused heavily towards preparation for the NMET. There is emphasis on grammar, reading comprehension, and vocabulary with students spending long hours on homework and exercises (Jin and Cortazzi, 2006).

The matriculation test battery as a whole comprises six subjects, with Chinese, Mathematics and English being compulsory subjects (Cheng, 2008). The matriculation test procedures are metaphorically labeled as a 'single-log-bridge', as for the overwhelming majority of students the results from this particular test are the sole route to university entrance. Various modifications have been made to the NMET since its inception in 1985, but it remains a standardized norm-referenced proficiency test

largely based on multiple-choice items (Qi, 2007). From students' perspectives, their focus is on achieving a satisfactory numerical score so they may engage in strategic behaviours which run counter to intended purposes of test-designers or teachers. For example, although test designers hope students would be encouraged to write for communicative purpose, students are mainly interested in aspects of writing that would help them achieve a better score and neglect the development of skills to write communicatively (Qi, 2010).

The study took place in a major city in Hubei province, chosen because of the first author's understanding of the context based on her previous experience as a teacher in the high school where this study is conducted. This school is among the top two key high schools in the city and only students with successful test performance in the municipal level entrance exams can enter it. Over 90% of students in the school are usually admitted to university, so a major competitive goal in the school is for admission to highly ranked universities.

Tests for the subject of English are held every other week in the school and this increases to weekly tests in the final year of high school as the NMET approaches. Students are ranked, within the class and school, and also take part in joint exams with neighboring schools to benchmark their performance against others. In addition to tests, students are assessed in a variety of other ways, including exercises, quizzes and a dictation conducted on a weekly basis to check students' memorization of the spelling and use of words. The so-called dictation is more like an exercise checking the mastery of words and phrases than the literal meaning of dictation (i.e. students writing down what is being read). Teachers usually read or present words and phrases they would like to assess, and ask students to do related exercises (for example, spelling, translation and blank filling).

## **Method**

This study addresses the following research question:

How do selected students at a high school in Hubei perceive their experience of English language assessment?

### *Participants*

The participants in the study were in the second year of a three year senior high school program, so the NMET would take place one year later. This group was chosen because they already have experience of high school assessment, whereas they are not in the highly pressurized final year when researching them might be less ethically desirable. Twenty-nine students (15 male, 14 female, all aged 16 or 17) spread evenly across three classes and different abilities agreed to participate.

### *Data Collection*

The main data collection methods were draw-a-picture technique and related interviews. Draw-a-picture technique involves students drawing pictures to illustrate aspects of their

school assessment experience. This technique has been found particularly useful in allowing children to share their personal feelings about schooling (Freeman and Mathison, 2009). A range of validity and reliability evidence has been collected (Haney, et al., 2004) which indicates that drawings document real and important aspects of student perceptions. This research strategy is increasingly being used to probe students' feelings about their experiences of assessment (Brown and Wang, 2011; Carless and Lam, 2012; Harris, et al., 2009).

The prompt and the procedures for the draw-a-picture method were piloted with two groups of students in the school. The main data collection was then conducted over a four week period of investigation. To facilitate the drawing procedure, a written prompt was provided (modified from Harris, et al., 2009) and students were invited to draw a picture of English assessment based on their own understandings, experiences and feelings, and also include an explanatory caption (later translated from Chinese to English). Students were then given 10-15 minutes to draw pictures; alternatively, some students chose to complete them during their free time. One student resisted the idea of drawing pictures and instead wrote sentences about her perceptions.

The first author then conducted individual or focus group interviews with students, taking their preferred type of interview into account. Altogether ten individual interviews, six paired interviews, and two group interviews involving three and four students respectively were conducted. The interviews started by asking students to describe their pictures and then follow-up questions encouraged participants to elaborate their views of both their pictures and wider experiences of school assessment. All interviews were conducted in Chinese to facilitate student expression of their perceptions and were then transcribed and translated into English.

### *Data Analysis*

Our coding of the data began with content analysis through which we counted the frequencies of various visual images in order to identify the main trends in the drawings. Content analysis is an approach to the analysis of visual or textual data that seeks to quantify content in a systematic manner in terms of developed categories (Bell, 2001). Based on this initial content analysis, we condensed the data into three major categories (see table 1).

First, affective responses were a major theme and we coded as 'Affective-negative', elements such as unhappy facial expressions, downturned mouths, bowed heads and captions indicating sadness related to assessment. Conversely, pictures and captions denoting, for example, satisfaction with or pride in high grades or teacher praise were coded as 'Affective-positive'. Second, we coded elements as 'Summative aspects' when the drawings illustrated student test-taking and/or the provision of summative results. Summative aspects were further divided into three sub-categories, including test scores and rankings (coded as 'summative aspects-test results'); rows of students sitting at separated desks or clocks ticking (coded as 'summative aspects-test procedures'); and pictures which suggested limitations of tests (coded as 'summative aspects-test limitations'). Third, data were coded as 'Formative aspects', when drawings suggest that assessment was used in ways which seemed to focus on trying to improve students' learning. These

**Table 1.** Classification of Findings from Drawings.

	No.	%
<b>Affective Responses</b>		
Negative	21	21.2
Positive	14	14.1
<b>Summative Aspects</b>		
Test results	18	18.2
Test procedures	12	12.1
Test limitations	4	4.0
<b>Formative Potential</b>		
Teacher support	14	14.1
Peer feedback	5	5.1
Student follow-up	7	7.1
Others	4	4.0
<b>Total</b>	<b>99</b>	<b>100</b>

include pictures of teacher feedback, peer support or intended follow-up, coded respectively, ‘formative potential-teacher support’, ‘peer feedback’ or ‘student follow-up’.

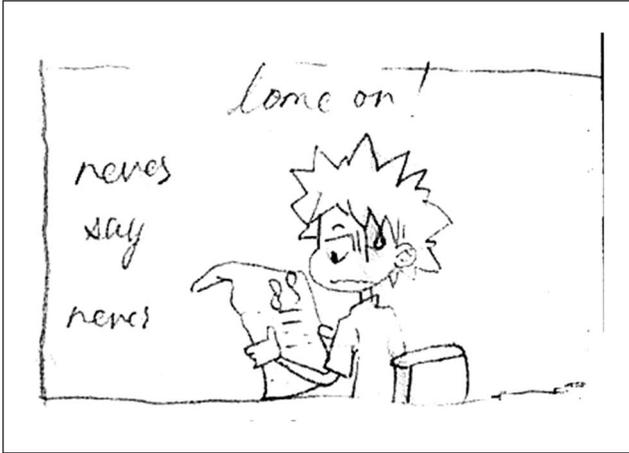
Some students chose to compose multiple small pictures to illustrate what they wanted to say and in total, the 29 students generated 68 pictures. We found that within a single picture, students also often expressed multiple elements and when this occurred we assigned two codes, where appropriate. For example, a student feeling happy because of advice or support from a classmate is coded as ‘Affective-positive’ and ‘Formative potential-peer feedback’. As many of the pictures expressed multiple meanings in this way, the number of codes (99) exceeds the number of drawings (68).

### *Limitations*

A number of limitations of the study need to be acknowledged. In line with its qualitative research design, the study is only from a single school in a province in China. There are also a relatively small number of student participants, although we believe the number is sufficient for our current exploratory purposes. The main limitations of drawings are that some students may lack drawing ability and drawings may be ambiguous. These issues are addressed in our study by requiring students to provide captions to explain their drawings; and more importantly through detailed follow-up interviews which carry both a triangulation function and allow students to expand on the meaning of their drawings.

### **Findings**

Table 1 classifies the drawings data by the three categories and eight main codes. The three categories of affective responses, summative aspects and formative potential are discussed in subsequent sub-sections. Pictures are interwoven with the text to represent the main findings and other pictures referred to by students are placed in an Appendix.



Caption: I hate exams and the marks do hurt me a lot (S1).

**Figure 1.** Student Affective Response to Exams.

### *Affective Responses*

Altogether 21 pictures were coded as indicating negative affective responses and 14 as positive.

*Negative Affect.* Negative affective responses of students were the most frequently occurring element in the data. Images of negative emotions include: tears, downturned mouths, perspiring faces or lowered heads. These symbols often occurred together with indicators of low marks, teachers' criticism or images of testing taking place. Below is a picture representing 'Affective-negative' (and it is also coded as 'Summative aspects-test results').

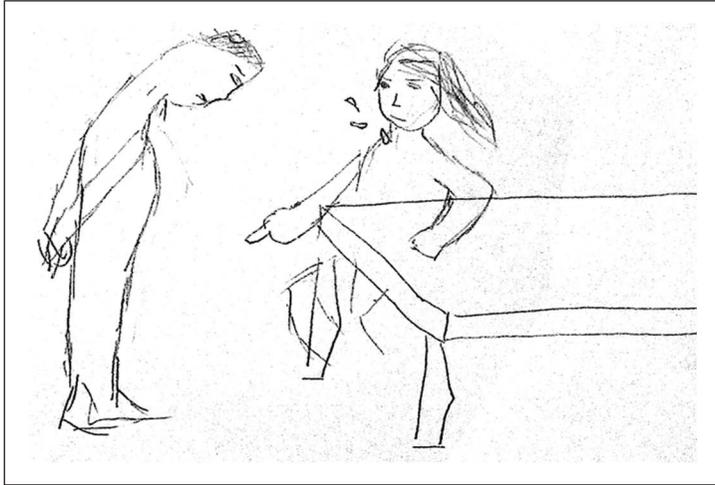
When asked to explain his picture, Student 1 (S1) responded:

Exams ruined students' lives. They took up our time, made us stressful and took away our happiness ... We need to face the NMET. Progress is just reflected in marks ... You have no chance to go to university without good marks and people judge whether you are a good person by the marks you get.

Despite these negative sentiments, there were also other perceptions expressed in the drawing. For example, 'never say never' and 'come on' seem to represent persistence and determination. Different dimensions of this student's response were also evident in other data which revealed that although he was negative about tests, he was positive towards teacher and peer feedback, and his own initiatives to reflect on and improve his learning.

A negative influence of tests on student life was depicted in another picture (Appendix 1, picture 1) with a caption reading:

When I was a child, there were no exams and life was like sunshine; it turned cloudy later because of exams. We have to memorize answers for exams and work day and night to finish endless homework (S2).



Caption: I felt anxious and nervous when the teacher talked to me (S3).

**Figure 2.** Student Affective Response to Teacher-Student Communication.

S2 explained that she was studying mainly to achieve success in the NMET and this restricted the scope of her English language learning. In her view, a more attractive way of learning English involves reading various books and communicating with others through English.

Negative emotions are also illustrated through nervousness in relation to teacher-student communication after tests. In their portrayal of conversations with teachers, participants drew images of students weeping or lowering their heads in shame. Figure 2 illustrates a student bowing his head with his hands clasped behind his back, when the teacher talks to him. It is coded as ‘Affective response-negative’.

S3 explained his feelings about communication with the teacher:

We all felt nervous. I only wanted to be called out when I got good marks, but this never happened. Teacher told me to spend more time in doing reading comprehension because it is important in the exam.

He added that the teacher’s suggestions were almost the same every time she spoke to him and added to the feelings of pressure he was experiencing.

Our interpretation of this data is that negative student affect mainly arises from the need to obtain high scores in the NMET. Related pressure from teachers also exacerbates nervousness and anxiety amongst some students. Impossible to gauge from our data was the extent to which this pressure was harmful to student well-being or instead prompted students to sustain their study efforts.

*Positive Affect.* Images of positive response to assessment relate mainly to student sense of achievement, receiving praise from teachers and support from peers or teachers. Figure 3 (coded as ‘Affective-positive’ and ‘Summative aspects-test results’) is an example of a drawing of a student who received praise from the teacher.



Caption: The praise from English teacher can motivate the student and his marks increase (S4).  
**Figure 3.** Student Affective Response to Teacher Praise.

The picture depicts an English teacher (ET) praising a student. The picture indicates a pattern of increasing marks from 105 to 135 (the total mark for this English test is 150). S4 reported that he makes more progress when receiving positive comments and encouragement from teachers, as these could motivate him to improve.

Our interpretation is that positive affect seems to arise mainly from achievement or perceptions of achievement. In a competitive examination system, we speculate that those who think they are likely to perform well develop positive affect, whereas students who believe they are unlikely to achieve their goals will feel less happy. Part of this may be due to individual factors, such as personality or confidence.

*Summary.* In sum, negative emotions come mainly from the stress of preparing for and taking tests, and nervousness or anxiety in relation to teacher-student communication about testing. Students mainly expressed positive feelings in relation to satisfaction from achievement and praise or encouragement from their teacher.

### Summative Aspects

Thirty-four pictures were coded as representing summative aspects of assessment: 18 within a theme of test scores and results, 12 referring to test procedures and 4 invoking student perceptions of various limitations of testing.

*Test Scores and Results.* Eighteen student pictures focused on issues related to marks, grades, test performance and rankings. Visual symbols included numerical marks, ticks, crosses or evaluative comments. S5 drew a picture entitled 'Ranking' (Appendix 1, picture 2) with a list of numbers on the left-hand side and names on the right, ranked by their marks. He commented that ranking by teachers of the student test scores is a

common practice and that this kind of ranking enables him to gauge his prospects for admission to his preferred university. S6 depicted a series of pictures (Appendix 1, picture 3) about test results, one showing he got 150 and the English teacher praised him, another showing an unhappy classmate who got zero. We infer that the frequent use of marks and rankings as indicators of students' academic performance was a social norm related to the competitive nature of the NMET.

*Test Procedures.* Summative aspects of assessment were also represented by 12 pictures of rules and regulations concerning testing procedures, such as rows of seats with desks placed apart; teachers acting as invigilators; or clocks on the wall showing limited time remaining. S7 depicted a student perspiring when the exam came to the end because of the race against the clock (Appendix 1, picture 4). The caption read, 'The time passes so fast. It is almost the end'. He also wrote a sentence from the blackboard: Be disciplined in exams. This discipline was not always obeyed as S8 drew himself copying his classmate's reading comprehension answers during a test (Appendix 1, picture 5).

*Test Limitations.* A third sub-theme in this category relates to four students' critiques of English tests, suggesting some of their limitations. In relation to figure 4, S9 explained that the top picture depicts her achieving a high mark in the English test (135 out of 150) and she felt satisfied. In the bottom picture, the question marks around her head illustrate that she felt puzzled when trying to communicate with a native speaker of English. She commented that she found both herself and her peers had difficulties in communicating in real life despite being able to obtain high marks in English tests.

A similar opinion was shared by S10, she drew an English test paper and a magazine article (Appendix 1, picture 6) and the caption was:

Doing test papers—is that useful? We do not know how to use English and only know how to deal with test papers (S10).

Both S9 and S10 appear to be indicating that the ability to complete test papers does not equate to skills required for using or understanding the language in real-life situations.

Another picture relevant to this category depicted a geometric object of irregular shape observed by four persons from four angles, resulting in four different conclusions about the shape (Appendix 1, picture 7). The caption was as follows:

Different conclusions can be drawn when a geometric object is observed from different angles. Assessment should be conducted from multiple angles, likewise English assessment. A single test cannot reflect our language abilities (S11).

S11 elaborated in his interview that multiple aspects of English abilities could include: spoken English, knowledge of vocabulary and motivation for learning.

Our interpretation of this data is that students are suggesting tensions between the ability to take tests and mastery of communicative English use. Preparing for written tests in the kinds of format required by NMET may not encourage the development of



Caption: Although we can do quite well in exams, we still find it difficult to communicate with foreigners. The purpose of learning a language is to use it, but we are learning English for exams (S9).

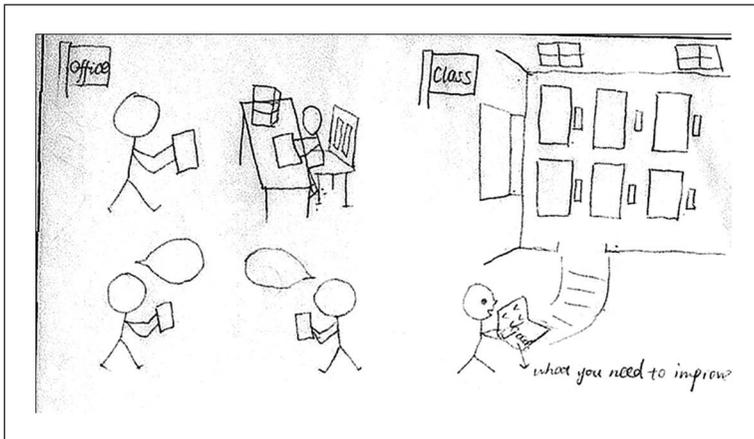
**Figure 4.** Test Limitations.

communicative language skills. This represents a challenge facing teachers and test developers in high-stakes contexts. The premium on reliability of scores and the use of multiple-choice testing may preclude or discourage forms of assessment which are focused on student production of communicative English.

*Summary.* In summary, this category has illustrated three themes: a focus on test scores and rankings; procedures and regulations of examinations; and perceptions of limitations of tests in encouraging use of communicative English.

### *Formative Potential*

We identified 26 instances where students expressed through their drawings what we interpreted as reference to formative potential of assessment. Of course, this does not imply that a picture represents formative assessment per se; we are instead suggesting that the student is identifying some process, such as feedback, which has potential to



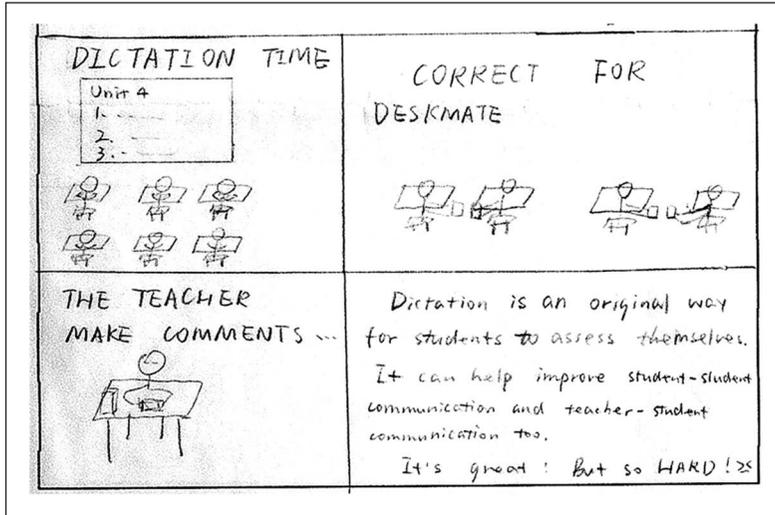
Caption: When I finished the exercises, the teacher checked my homework, made comments and communicated with me face to face. Afterwards, I am more aware of what I need to improve most and I will be motivated for future study (S12).

**Figure 5.** Teacher Feedback and Intended Student Follow-up.

support the development of their learning. These processes emanated from teachers (14 times), peers (5 times) or student follow-up actions that they envisage (7 occurrences), discussed in three subsequent sub-sections.

*Teacher Support.* Student pictures identify 14 instances of support from teachers in terms of teacher-student communication and teacher feedback. In figure 5, the picture on the left (coded as ‘Formative potential-teacher support’) represents a meeting between teacher and student (S12) in the teacher’s office. The picture on the right illustrates the student returning to class with a smile on her face. This was assigned two codes: ‘Affective-positive’ (derived from her statement of motivation and positive facial expression) and ‘Formative potential-student follow-up’ (her awareness of what she needs to improve as indicated in the caption). S12 reported that she received some encouragement from the teacher which makes her feel motivated and that the identification of weaknesses assists her in becoming clearer about future goals.

Other students, for example S13, expressed more mixed feelings towards teacher-student communication. Her picture shows a teacher talking to her with a downturned mouth and upturned eyebrows when her performance in dictation was unsatisfactory (Appendix 1, picture 8). She stated in her interview, ‘the teacher was harsh and she pointed out my problem directly. She criticized me for not devoting sufficient time to memorization and revision. It sounded unpleasant, but it was helpful. I did not take it seriously when my previous teachers pointed problems out in a less direct way’. We infer that S13’s response represents a tension between frank and more tactful feedback. Direct criticism may help students form a realistic understanding of their standard and what they need to improve, but risks being discouraging. Indirect and less forceful feedback may be more palatable to students, but may not stimulate improvement. Usefulness of



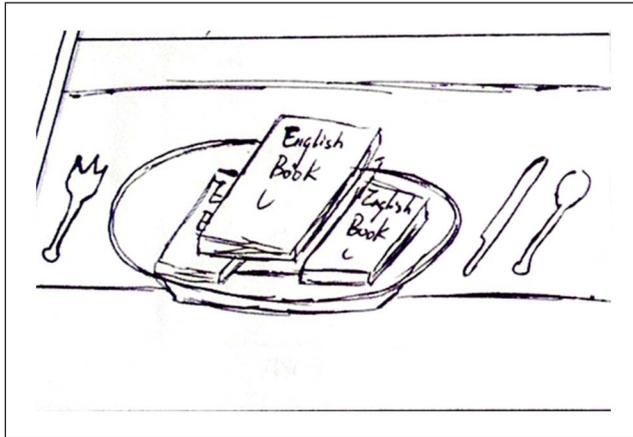
**Figure 6.** Peer Feedback in Dictation.

different forms of feedback is likely to depend on the individual characteristics of students, such as personality, confidence and attitude.

*Peer Feedback.* Four pictures describe peer dictation which the classroom observations showed to be a process through which teachers ask students to correct their desk-mates' work and write comments for them. Figure 6 drawn by S14 depicts the peer dictation process through a sequence of three pictures (and a caption, bottom right) which were coded as 'Summative aspects-test procedures' (top left), 'Formative potential-peer feedback' (top right) and 'Formative potential-teacher support' (bottom left).

S14 commented that this kind of peer feedback promotes emotional support, especially when peers have good relationships with each other. Although S14 regards the peer support as positive, she also expressed the view that it does not provide particularly useful academic advice because the comments are added more to complete the task than to provide insight. S15 drew a picture of herself holding her friend's hand (Appendix 1, picture 9). She explained that her classmate supported her, for example, through checking her memorization and understandings of words and phrases before dictation. Other students, however, perceive teachers' comments are more helpful for academic improvement and peer feedback can become repetitive as there is a limited amount of advice peers can provide e.g. 'be more careful'. We interpret that the potential for useful student feedback also relates to the nature of the assessment task, for example, checking classmates' reproduction of vocabulary is likely to lead to limited feedback messages. Overall, the small amount of evidence on this theme indicated that peer feedback seemed to be mainly useful for encouragement and support, with advice being modest in substantive value.

*Student Follow-up.* Student follow-up after a test was portrayed through seven pictures with figure 7 being an example. This picture (coded as 'Formative potential-student



Caption: I got a lot wrong in dictation. Textbooks are important and I need to digest them (S16).

**Figure 7.** Student Follow-up.

follow-up’) depicts how student S16 reacts to his unsuccessful performance in a dictation. He uses the image of a dish of English books to be digested and explained as follows: ‘I found that all things dictated come from textbooks so I need to have a better command of what is in them’.

Other pictures demonstrated efforts made in test follow-up, including S17’s image of perspiration while memorizing irregular verbs (Appendix 1, picture 10). When asked to elaborate, S17 explained, ‘The teacher criticized me for not studying attentively, so I tried to memorize all she has taught in class. My performance in dictation became better this time’.

A feature of students’ reported follow-up actions is that they seemed to focus mainly on test-related exercises, for example, doing reading comprehension exercises or multiple-choice items for practising vocabulary and grammar. Students also articulated that they experienced difficulties in implementing follow-up actions, for example, there was new content to learn which left little time for follow-up.

Our inference is that student follow-up (if carried out fruitfully) carries potential to support a formative orientation within the process of summative assessment. The issue of the extent to which their follow-up efforts represented genuine long-term learning or a more limited short-term memorization to boost marks was an issue which is difficult to resolve.

**Summary.** In short, formative potential was identified through three strategies: teacher support or feedback; peer feedback; and student follow-up actions.

## Discussion

The current study seeks to identify perceptions of English language assessment of a small group of Chinese high school students. We found that students’ negative perceptions of

assessment exceeded positive ones by a margin of 21 to 14. Negative feelings mainly arise from three aspects: pressurized school life and heavy workload; perceived negative sentiments following from unsatisfactory or disappointing results; and discouraging conversations with teachers. It was also noteworthy that there were a number of positive reactions to assessment. Positive affect mainly arises from satisfaction with attainment, such as high marks and teacher praise, and feelings of being supported or encouraged by teachers and peers.

Returning to the framework for the paper, our data have reinforced the power of testing in the lived experience of students. The student responses come from a particular test-dominated setting and reinforce the notion that assessment is a highly contextualized phenomenon influenced by various socio-cultural factors. In China, these may include: the long history of competitive examinations; the large candidatures which may accentuate competition; and societal emphases on academic achievement as a measure of an individual's worth. It is possible, for example, that students educated in such contexts develop stoicism to face the challenges of sustained pressure to succeed in tests and examinations.

Our research bears comparison with that of Huhta et al., (2006) in Finland, in that the students in both studies are of a similar age and are preparing for a high-stakes matriculation examination. A particular strength of the Huhta et al., study is that it traced student responses to assessment over a four-month period, whereas our study only captured their views at a particular point in time. Both studies reinforce the emotional impact of assessment, with a difference being that some of the Finnish students seemed to possess more confidence towards testing than their Chinese counterparts in our study. It is possible that relatively direct (or even harsh) feedback in the Chinese context is a contributing factor and there may be cross-cultural implications in relation to the provision of feedback to students on their assessment performance.

Recent literature (Carless, 2011; Davison and Leung, 2009) has mainly discussed the relationship between formative and summative assessment from a teacher perspective and our study adds students' perceptions of how they interpret assessment messages. Summative indicators seemed more prominent in that examinations were viewed as providing marks and rankings which impacted on how students perceived themselves. A number of students, however, also noted potential in the processes surrounding testing to provide information which could be used for improvement purposes, or at least they seemed to be aware of the necessity of taking follow-up actions. These were targeted mainly at improved test performance rather than wider views of communicative language use or learning for its own sake. An enduring challenge for the NMET (and other test development) relates to the tension in fulfilling needs for reliability in high-stakes testing as well as promoting wider communicative competence through positive washback (cf. Qi, 2007, 2010).

What the paper has done differently to previous ELT research is in using draw-a-picture technique as the starting point of student expression of their perceptions and triangulating it with interview data. Drawing pictures is a relatively student-centered research strategy which allows students to express their thoughts in their own way with less constraint than that imposed by interview protocols. An additional nuance is that drawing is individual work and so less likely to be influenced by group conformity than focus group discussions. The preparation of pictures prior to an interview also provides students with

an initial opportunity to share their thoughts and can alleviate the difficulty of ‘breaking the ice’ at the beginning of an interview. It is also our contention that draw-a-picture technique is an effective way of exploring emotionally charged issues, such as assessment. ELT researchers might consider further drawing-based studies with students in different contexts or with different age groups. The recent paper (Carless and Lam, 2012) which involved students aged six to eight years old suggests that drawing pictures might be particularly suitable with young learners, although it has been used in recent educational assessment research with other age groups (e.g. Harris et al., 2009; Brown and Wang, 2011). An issue for further exploration might be the extent to which there might be cultural differences impacting on the usefulness of pictures as opposed to interviews in facilitating free student expression of their innermost feelings. Is draw a picture technique more meaningful in certain contexts than in others?

## Conclusion

This study has discussed perceptions of assessment from a group of Chinese high school students and has revealed some of the different dimensions of student responses. Although the findings may reveal relatively few surprises, we believe that documenting students’ perceptions in some detail makes a valuable contribution to the knowledge base and can act as a springboard for future research efforts. This study has also explored the possibilities of draw-a-picture technique as a means of eliciting students’ perceptions and suggested that it might merit further use in ELT research. Finally, we suggest some further possibilities for language assessment research in China and elsewhere. How do students in China perceive the relationship between the high-stakes NMET and classroom assessment? To what extent do these relationships act as a stimulus or hindrance to student cognitive engagement? What are effective ways of deriving meaningful language learning from classroom processes dominated by summative assessment? Such issues at the interface of summative and formative assessment also have wider import for ELT research (cf. Davison and Leung, 2009).

## Funding

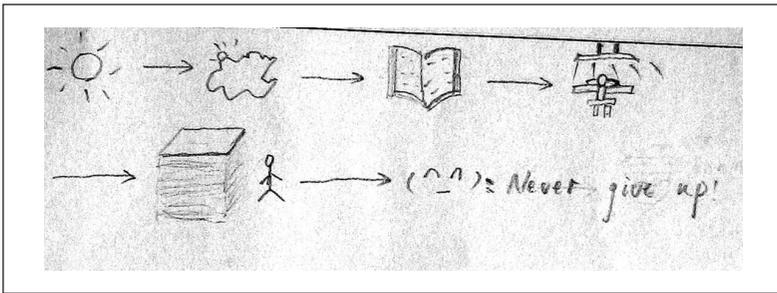
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

- Bell P (2001) Content analysis of visual images. In: van Leeuwen T, Jewitt C (eds) *Handbook of Visual Analysis*. London: SAGE, 10–34.
- Brown GTL, Wang Z (2011) Illustrating assessment: how Hong Kong university students conceive of assessment’s purposes. *Studies in Higher Education*. DOI:10.1080/03075079.2011.616955
- Butler YG (2009) How do teachers observe and evaluate elementary school students’ foreign language performance? A case study from South Korea. *TESOL Quarterly* 43(3): 417–44.
- Carless D (2011) *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-heritage Settings*. New York: Routledge.
- Carless D, Lam R (2012) The examined life: perspectives of lower primary school students in Hong Kong. *Education 3–13: International Journal of Primary, Elementary and Early Years Education*. DOI:10.1080/03004279.2012.689988

- Cheng L (2008) The key to success: English language testing in China. *Language Testing* 25(1): 15–37.
- Cheng L, Qi L (2006) Description and examination of the National Matriculation English Test. *Language Assessment Quarterly* 3(1): 53–70.
- Davison C (2007) Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly* 4(1): 37–68.
- Davison C, Leung C (2009) Current issues in English language teacher-based assessment. *TESOL Quarterly* 43(3): 393–415.
- Deng C, Carless D (2010) Examination preparation or effective teaching: conflicting priorities in the implementation of a pedagogic innovation. *Language Assessment Quarterly* 7(4): 285–302.
- Dörnyei Z, Ushioda E (2011) *Teaching and Researching Motivation (2nd edn)*. Harlow: Longman.
- Freeman M, Mathison S (2009) *Researching Children's Experiences*. New York: Guilford Press.
- Haney W, Russell M, and Bebell D (2004) Drawing on education: using drawings to document schooling and support change. *Harvard Educational Review* 74(3): 241–72.
- Harris LR, Harnett JA, Brown GTL (2009) 'Drawing' out student conceptions: Using pupils' pictures to examine their conceptions of assessment. In: McInerney DM, Brown GTL, Liem GAD (eds) *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning*. Charlotte, NC: Information Age Publishing, 53–83.
- Huhta A, Kalaja P, and Pitkänen-Huhta A (2006) Discursive construction of a high-stakes test: the many faces of a test-taker. *Language Testing* 23(3): 326–50.
- Jin L, Cortazzi M (2006) Changing practices in Chinese cultures of learning. *Language, Culture and Curriculum* 19(1): 5–20.
- Leung C, Lewkowicz J (2006) Expanding horizons and unresolved conundrums: language testing and assessment. *TESOL Quarterly* 40(1): 211–34.
- Leung C, Rea-Dickins P (2007) Teacher assessment as policy instrument: contradictions and capacities. *Language Assessment Quarterly* 4(1): 6–36.
- Li M, Baldauf R (2011) Beyond the curriculum: a Chinese example of issues constraining effective English language teaching. *TESOL Quarterly* 45(4): 793–803.
- Oxford R, Shearin J (1994) Language learning motivation: expanding the theoretical framework. *The Modern Language Journal* 78(1): 12–28.
- Qi L (2007) Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy and Practice* 14(1): 51–74.
- Qi L (2010) Should proofreading go? Examining the section function and washback of the proof-reading sub-test in the National Matriculation English Test. In: Cheng L, Curtis A (eds) *English Language Assessment and the Chinese Learner*. London: Routledge, 219–33.
- Shohamy E (1998) Critical language testing and beyond. *Studies in Educational Evaluation* 24(4): 331–45.
- Shohamy E (2007) The power of language, the power of English language and the role of ELT. In: Cummins J, Davison C (eds) *International Handbook of English Language Teaching, Vol. 15*. New York: Springer, 521–31.
- Shohamy E, Donitsa-Schmidt S, Ferman I (1996) Test impact revisited: washback effect over time. *Language Testing* 13(3): 298–317.
- Wang Q (2007) The national curriculum changes and their effects on English language teaching in the People's Republic of China. In: Cummins J, Davison C (eds) *International Handbook of English Language Teaching Vol. 15*. New York: Springer, 87–105.
- Xiao Y, Sharpling G, and Liu H (2011) Washback of National Matriculation English Test on students' learning in Chinese secondary school context. *Asian EFL Journal* 13(3): 103–29.
- 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.

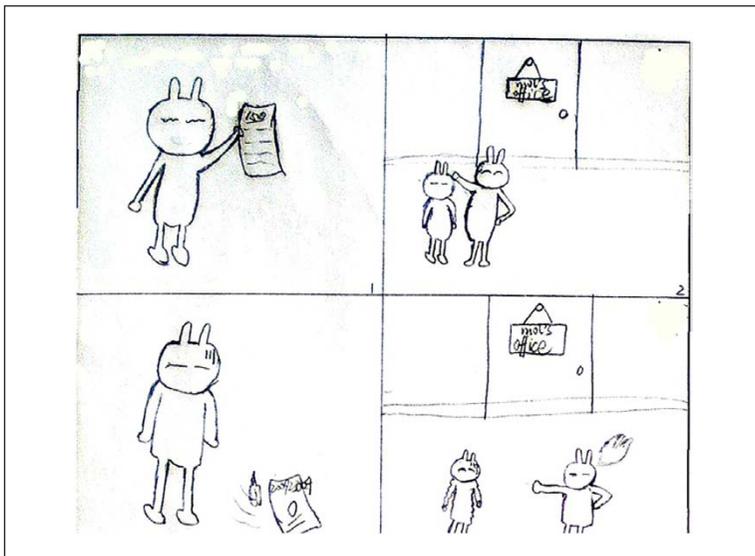
Appendix I



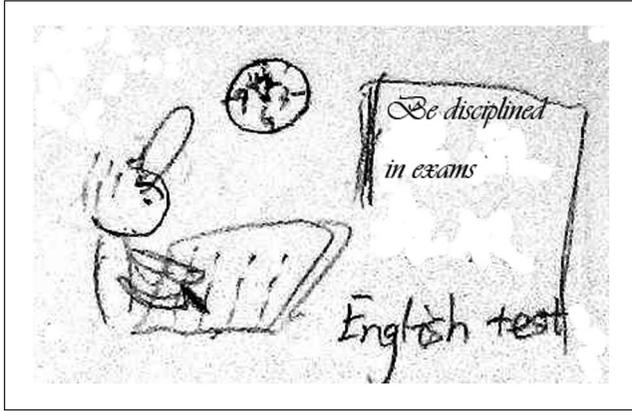
Picture 1. (S2)



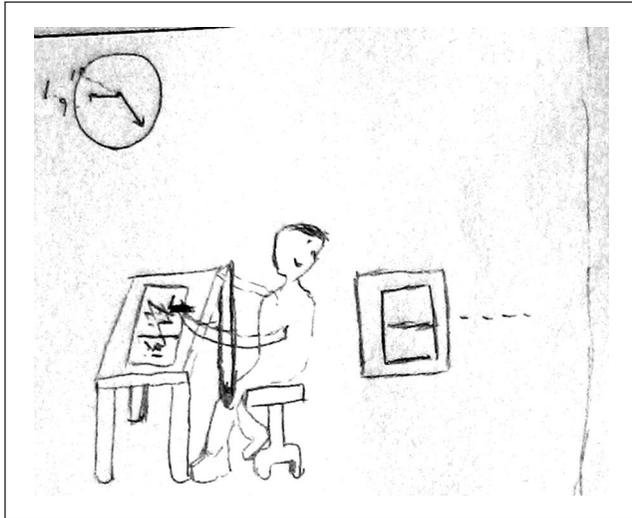
Picture 2. Caption: There are rankings and cut-off scores for universities and key universities (S5).



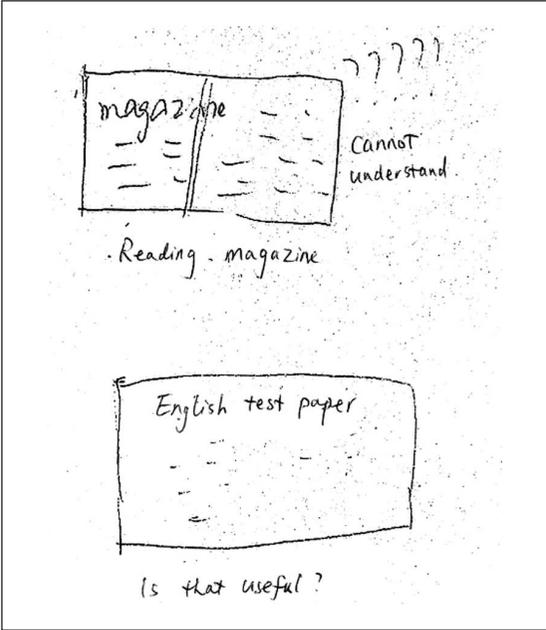
Picture 3. Caption: I got good mark, the teacher praised me. A student got zero, the teacher shouted at him (S6).



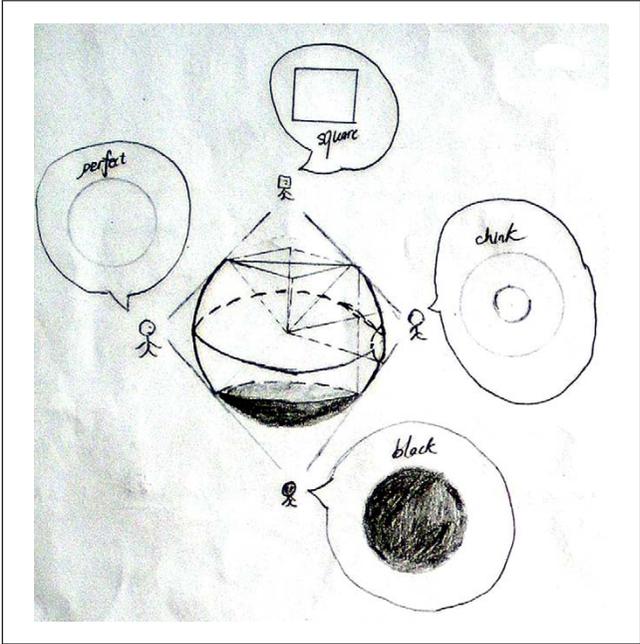
Picture 4. (S7)



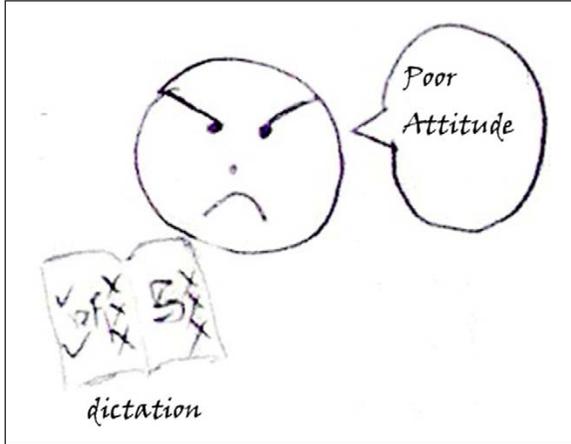
Picture 5. Caption: I do not have patience to do reading comprehension and always copy my friend's (S8).



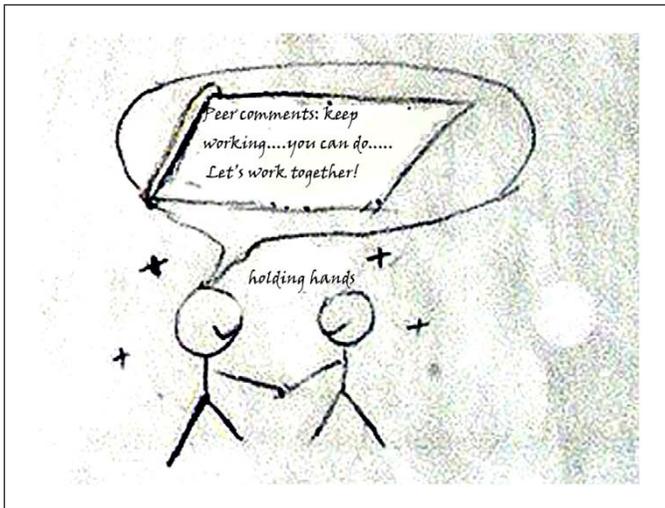
Picture 6. (S10)



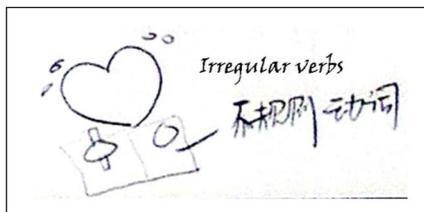
Picture 7. (S11)



**Picture 8.** Caption: The teacher criticized me for the poor performance in dictation (S13).



**Picture 9.** Caption: We give each other peer feedback. We find out problems and make improvement together (S15).



**Picture 10.** Caption: I worked hard to memorize irregular verbs and put my heart into it (S17).