CHO06106

Assessing Taiwanese Students’ English ability:
Semantic Competence versus Pragmatic Competence

A paper to be presented at the

Australian Association of Educational Research
2006 Annual Conference, Adelaide
27th-30th November

By

Tungshan Chou
National Hualien University of Education, Taiwan

Gordon Woodbine
Curtin University of Technology, Australia

&

Hiewu Su
National Donhwa University, Taiwan
Abstract

The problem of current English assessment in Taiwanese classroom settings was visited. A recommendation to assess English ability in the notion of semantic and pragmatic competence was made. An instrument was constructed in line with our recommendation and administered to 341 middle school and college students. We hypothesized that in terms of item percent correct, items measuring both semantic and pragmatic competence will become easier as students progress from middle school to college; whereas in terms of item discrimination, semantic items are more discriminating than pragmatic items in the current context. Empirical analyses confirmed our hypothesis and the implication discussed.
Background

The issue of “pragmatic English ability” and how it may be assessed in the learning of English as a foreign language in the Taiwanese context has for a long time been a strong interest of ours. We strongly suspect that the English education throughout all levels of formal schooling in Taiwan is failing to develop students’ functional abilities in using English as a foreign language. Though neither of the authors actually came from the field of teaching English as a foreign language, we decided to embark on this research, both as informed observers of current teaching practices and as concerned parents of current school-age children, in the hope that some light may be shed in regard to uncovering why the system fails to produce functional students. We chose the area of formal assessment as our stepping stone into understanding how the English education operates in Taiwan nowadays.

Our interest was initially triggered by the contents of some major English examinations our then middle-school age daughter brought home periodically. From these tests we were often struck that the abilities required of students to score high do not necessarily equate with the ability to function in an ordinary daily living English language setting. After having spent a few years of her primary school education in an English speaking country, she found that English taught in her Taiwanese school was not the same English language she was familiar with. She had to adjust herself in many ways in order to fit into the system. Some of the major differences she noted were in pronunciation, inflexible choices of words for objects and ideas, and the conception of “correct” English grammar, and unfortunately these differences often constitute a significant proportion of the day-to-day assessment activities in English classrooms.

Our interest was reinforced when we were solicited by the parents of the students in her form class to serve as the teacher’s assistants to help her maintain classroom order on a one-hour per week basis. While serving in this role we chatted with mainstream English teachers who had students approximately the same age as ours (then 14-15). We mentioned our concern in regard to both the methods of teaching and the contents of assessment, but to our surprise our suggestion to bring the contents of English education closer in line with developing students’ pragmatic English ability was met largely with strong resistance. To their credit indeed, most of these teachers do teach in the belief that they teach in ways that are most beneficial for their students in the Taiwanese environment. However, we did find solace in some parents in the course of working with the school who shared our views. Most of these
parents are not grass-root kind of people. Rather, they are well educated and mostly have had overseas experiences.

Types of Items Incompatible with Developing Students Pragmatic Ability

With regard to pronunciation, a lot of effort is devoted to linking English sounds with similar sounds in Chinese, such as the letter “c”, which is commonly taught to be pronounced as the Chinese word “shi, 西”, which would most likely be incomprehensible to most native English speakers. Similarly, the letter “n” is often likened to the Chinese word “恩”. The letter “w” is often pronounced with stress placed on the last syllable instead of the first. The distorted version of pronunciation by itself would not pose a serious problem in our opinion, as a special accent is often associated with a particular nation or culture; i.e., “thanks” is often pronounced as “tanks” in Indian and associated cultures. The real serious problem is the lack of knowledge that certain words can be pronounced differently in different parts of the world. Take the following item for instance:

Which of the following has the same pronunciation as in data?
(1) Table  (2) Pat  (3) Mat  (4) Gather  (5) Cat

The school provided correct answer is (1), according to the North American way of pronunciation. But in fact, all the remaining four options are also correct, if you come from England or the British Commonwealth countries such as Australia. To say the other options are wrong is a blatant rejection of the diversity our world is showing us today.

What bothers us most is the type of items that require inflexible choices of words for objects or expressions. It also happens to be the most common type of ridiculous items in our observation. We were struck by this item that appeared in my daughter’s first major English exam:

How are you?
(1) Good.
(2) Fine.
(3) Fine, thank you.
(4) Fine, thank you. And you?
The correct answer was given as (4), because apparently it was copied from a conversation from the standard textbook. Other options are deemed incorrect simply because they are not identical to what the textbook says. Another common form of inflexibility is the choice of prepositions. Instead of teaching students the meanings of prepositions, many teachers present prepositions in certain combinations of words and only these combinations are deemed as correct answers in a test. Consider this item:

I am responsible _____ you.
(1) for
(2) of
(3) with
(4) to

The answer that the teacher is looking for is (1), but actually (4) makes perfect sense too albeit it takes on the opposite meaning. It is difficult to convince Taiwanese teachers that items of this sort with only one sentence can take on more than one correct answer. Another example is, while “It is very kind of you” is being wholeheartedly accepted as a legal sentence, “I kind of like you” is a no-no according to most English teachers because the preposition “of” is taught to be followed by a noun but not by a verb in its original form.

The type of items that I found most amusing, but certainly not conducive to developing students’ pragmatic English ability is probably the kind of items that are considered to be grammatically correct, whereas in fact they would most likely confuse even the native English speakers. This can be best illustrated by a multiple choice item found in one of my daughter’s major exams during her last year:

“Both John and Mary are not happy.” means:
(1) John is not happy.
(2) Mary is not happy.
(3) John and Mary are not happy.
(4) One of them is happy.

The correct answer is (4) according to the strict grammatical rule, but to most people who are familiar with the English language, (3) appears to be the most appropriate answer, albeit it is an awkward form of expression. Items of this sort abound in school English tests across Taiwan. It is certainly no surprise that students who are taught to
focus on grammatical structure of sentences cannot produce satisfactory results when real-life English language competency is required. This phenomenon seems to go contrary to the worldwide trend in teaching English as a foreign language: The integration between teaching and assessment should focus on the “language as a tool for communication rather than on language knowledge as an end in itself” (Brindley, 1995, p.158).

Government’s Effort in Promoting Communicative Ability

We make it no secret that we are proponents of communicative ability in regard to teaching English and assessing English ability as a foreign language. We argue that learning a language is to learn how to communicate in that language. Indeed our call for attention to students’ communicative language competence has been echoed by the actions on the part of the Government (China Post, 2003, p.4). To address the need of assessing students’ communicative competence in English language, a nationwide proficiency certification examination board was set up to grade students’ English ability into twelve levels according to their communicative competence in English. In an effort to address the problem that many local English teachers are poorly trained in communicative language ability themselves, the Ministry of Education also introduced a scheme of hiring 1,000 qualified English teachers from native English-speaking countries to serve as seed teachers at elementary and high schools. These seed teachers shall function as resource people who provide local teachers and students with training programs leading to communicative proficiency.

Despite the efforts at the governmental level, the school-based English tests continue to adopt largely the traditional contents of the past, not much attention has been given to the communicative aspect. In fact it is quite ironic that despite much of the contents of English textbooks at the primary and secondary school levels is communication-based, the concept of assessing communicative ability does not form the backbone of the assessment of English ability in schools. The current study was motivated by the curiosity to understand the current level of Taiwanese middle school students’ English communicative ability. We chose to adopt the traditional testing format in this research to assess students’ communicative ability for obvious reasons (minimal confusion for students, easy administration and scoring, and straightforward analysis results). The traditional testing format refers to the dichotomously scored item format familiar to all students such as multiple-choice and fill-in-the-blank types of questions. The current study represents an attempt to break away in contents from traditional exam formats, and adopt the essence of communicative ability in the
design of test items for assessment.

Assessing Communicative Ability

It is important to point out that the designs for assessment of communicative ability differ from the traditional test designs in two major features: it is situation specific and it is production oriented. The former refers to the fact that by the definition of communication the contents of assessment are specific to a certain audience (people of certain occupations, areas, etc). The latter refers to the need that the communication requires oral or written performance. The production orientation feature of communicative language testing cannot be measured by traditional dichotomously scored item formats, and thus will not be a concern in the test design for this study. Instead, our current study will suffice by investigating the situation specific feature.

By excluding the production orientation feature from our test design, we by no means downplay its importance in the assessment of communicative language ability, of which production is an integral part. On the contrary, we uphold its importance, and the paucity of test designs aiming to measure this part of communicative ability calls for more research into coming up with valid assessment instruments for measuring students’ abilities in this regard. Nevertheless, this endeavor is beyond the scope of our current study. In this study, we chose to use two types of competence to represent communicative ability: semantic competence and pragmatic competence. Our conceptualization was taken in part from Bachman’s (1991) framework for communicative ability.

Semantic Competence versus Pragmatic Competence

Semantic competence comprises those abilities involved in organizing the structure of language for recognizing grammatically correct sentences and understanding the meaning of words. This competence can be further broken into two types: grammatical and textual. The grammatical competence consists of knowledge such as vocabulary and syntax, whereas textual competence consists of knowledge of the conventions for joining words together to form a meaningful text according to rules of cohesion and rhetorical organization. Items designed in traditional testing formats are mainly aimed to measure this type of competence.

Pragmatic competence pertains to understanding linguistic signals used in
communication and how they are used to refer to persons, objects, ideas, and feelings. The concept of pragmatic competence includes illocutionary competence and sociolinguistic competence. The illocutionary competence refers to the ability to perform a function by means of speaking (e.g., asking someone to leave so that you may get rest). The sociolinguistic competence is “the sensitivity to, or control of the conventions of language use that are determined by the features of the specific language use context; it enables us to perform language functions in ways that are appropriate.” (Bachman, 1991, p.94) In other words, sociolinguistic competence refers to sensitivities to differences in naturalness and cultural registers in a communication context. This type of competence is often overlooked in the assessment of communicative ability.

The Current Study and Research Hypotheses

Pragmatics is the branch of linguistics which studies those aspects of language meaning which cannot be predicted from linguistic knowledge alone. An American philosopher Paul Grice, referred to by linguist Jean Aitchison (1999, p.97) as the “father of pragmatics”, provided a philosophical basis for assessing pragmatic competence in a second language (L2) discourse setting. Grice emphasized four rules of a conversation that help human beings communicate efficiently to one another: quantity, quality, relevance and manner. For the purposes of this study, the description offered by Grice has been adopted for assessing pragmatic competence in the setting of learning English as a foreign language. We do recognize that although pragmatic competence as a whole can not be effectively tested using a written testing format, there are aspects of language that can be tested for that give a good indication of the progress language users have made towards achieving pragmatic competence. Grice supplied a good starting point from which we can test students’ pragmatic competence by testing on the additional information given in a discourse beyond what can be explained by linguistic knowledge alone, namely, by applying the four rules of conversation.

The unique dimensionality of pragmatic competence in the assessment of communicative ability for students who learn English as a foreign language and its associated psychometric properties have been confirmed in a most recent study (Chou & Chen, 2005). They showed that items can be designed to measure the pragmatic aspect of the English communicative ability with good reliability and discrimination. In this study we seek to gain some understanding about the general level of pragmatic competence of the Taiwanese students in comparison with their semantic competence
in the use of English language. To get a better coverage of the whole age spectrum, we sampled students from two levels of schools: middle school (aged from 13 to 15) and college (aged from 19 to 22). An instrument designed to measure both semantic competence and pragmatic competence will be constructed and then administered to these students. The statistics (percent correct and discrimination) of all the items making up the instrument will form the basis for testing our hypotheses of interest. The following two major research hypotheses are then postulated:

1. With respect to percent correct, we hypothesize that the semantic items will show higher percent correct than the pragmatic items because students are more familiar with this type of items than with the other kind of items. Of course, if the English education has done its job right, items measuring both semantic and pragmatic competence will become easier as students progress from Year 7 of the middle school to college.

2. With respect to item discrimination, semantic competence will be more discriminating than pragmatic competence in the current context of the Taiwanese English education.

Method

Instrumentation

Three settings were given as three sections in the form of three mutually independent work samples from which students must find answers to a number of questions asked according to information given. The first two sections are related to pragmatic competence in which the correct answers are not given in the text explicitly, but instead, one must apply the four rules in reaching a correct answer. The third section measures semantic competence, whose answers can be found directly in the text. All items take on the form of multiple choice format, in which there is only one most appropriate answer out of four or five given options. The test was officially administered in two stages. In the first stage, a test of 19 items in the form of three work samples was given to students from two large metropolitan middle schools. In the second stage, an additional item was added to the third section measuring semantic competence, and the test was then given to some college students. The additional item was added to make the best use of space, as well as to bring the total number of items to a more familiar number, 20 instead of 19, following the advice of
a colleague who reviewed the results from the first stage data analysis.

The three work samples and their associated items are summarized in Figure 1. The option with an asterisk (*) is the correct answer.

Figure 1.

Test items in the form of three work samples

Work Sample One

‘Why can’t I go with you?’ asked Annie as her mom got ready to leave.
Her mom kissed her on the cheek, “Because you’re only six and children are not
supposed to go to meetings.”
“But how come Johnnie gets to go? He’s only five!” protested Annie.
“I’m sure his mom will drop him off at his uncle’s place,” Her mom explained, “if he
goes, he would make a racket.”
Annie said, “I won’t make a noise.”
“I’m sure you won’t,” agreed her mom, “but it will be a long meeting and not
interesting for you.”
Annie couldn’t think why a meeting about having a swimming pool in her
neighborhood would not be interesting.
“Why not?” asked Annie. “I love to swim.”
“We’ll only be talking about what to do,” explained her mom. “There won’t be a
swimming pool there.”
Figured 1 - Continued

1. Which of the following is true?
   (1) *A swimming pool is being planned.
   (2) A swimming pool is being built.
   (3) A swimming pool has been built for quite some time.
   (4) A swimming pool has just been built.
   (5) No swimming pool will be built.

2. "He would make a racket", racket means:
   (1)pain (2)toy (3)*noise (4)paddle (5)joke

3. What is the best logical sequence for the following three sentences?
   a. Mom tries to talk Annie out of going to the meeting with her.
   b. Mom is going to a meeting.
   c. Annie would like to go to the meeting.
   (1)a-b-c (2)c-b-a (3)c-a-b (4)*b-c-a (5)b-a-c

4. How many people are having a conversation in the paragraph?
   (1)six (2)five (3)four (4)three (5)*two

5. What can Annie say to be more persuasive?
   (1) Swimming is a good sport, I like to go swimming.
   (2) *If I am bored, I can read a book by myself.
   (3) I am tall enough to stand in the pool.
   (4) I am older than Johnnie. If he gets to go, I sure can go too.
   (5) I’ll go to bed early if I swim.

**Work Sample Two**

Peter: Well, that’s all from Peter Lin. It’s time to join Alex in the studio.
Alex: Yes, and this morning we are going to learn about dumplings.
Mary: I can’t wait.
Alex: Let’s start with the skin. You can buy them at your local market. You don’t want them too thick or too thin.
Mary: What do you put inside them?
Alex: Traditionally, there are two main recipes. One is pork and leek and the other is pork and cabbage. My favorite is pork and leek but many people prefer pork and cabbage. It’s a matter of taste.

Mary: Both of them sound absolutely delicious.

Alex: You must seal it well. Bring the edges together and pinch them together with your index finger and thumb.

Mary: What happens if I don’t do it properly?

Alex: They will fall apart when you cook them.

Mary: It sounds like a challenge. Let’s get started.

6. Where is this taking place?
   (1) In a school   (2) At home   (3)* On a TV show   (4) In a church.

7. What can we say?
   (1) * Alex can make delicious dumplings.
   (2) Mary knows how to cook dumplings.
   (3) Mary and Alex are married.
   (4) Peter Lin is teaching Alex how to make dumplings.

8. What do you need to make dumplings?
   (1) * Leeks and pork   (2) Pork and rice
   (3) Pork and carrots   (4) Beef, leeks, cabbage

9. Do you think?
   (1) Mary stays up late to learn about dumplings.
   (2) * She must get up early to learn about making dumplings.
   (3) Mary must get up early to buy skins.
   (4) Peter must get up early to buy cabbage.

10. What does this conversation encourage?
    (1) * Eat more dumplings.
    (2) People should make their own skins.
    (3) Talk to your friends more.
    (4) When making dumplings, buy the skins.

11. What is Mary waiting for:
    (1) To go home.       (2)* To start cooking.
    (3) To go to the bathroom.   (4) For Alex to stop talking.
12. Dumplings are made from _________.
   (1) A foreign recipe.       (2) A new fad.
   (3) A recipe passed down from past generations. (4) A great talking point.

13. “It’s a matter of taste.” means:
   (1) Everyone likes the same kind of food.
   (2) Only experts know what a good dumpling is.
   (3) *Each person has his/her own preference.
   (4) Dumplings taste good.

**Situation Three**

If you like reading, the Book Fair that’s being held this month will certainly interest you. The show is taking place between the 17th and 19th of March at the Cultural Center. It aims to raise awareness of the enjoyment of reading for all ages.

We’ve got lot’s of different things for you to see and hear. There will be stands showing books of all kinds for all ages, and talks by educational speakers on the best ways to encourage people to read. You will also be able to see an exhibition of classic books and their authors. We’re sure you will enjoy seeing demonstrations of latest book designs and meeting real live authors.

This is your chance to make good decisions about what to buy. You can attend the fair between nine-thirty and five on Thursday and Friday, and from nine-thirty to four on Saturday.

Tickets cost $500 each, or $250 if you are a full-time student. All tickets can be booked by ringing the hotline: (03) 8786789. Parking can be really difficult around the Cultural Center. However, an all day parking space can be booked by calling the hotline: (03) 8782222.

We look forward to seeing you there.

14. When will the book fair take place?
   (1) 17th March       (2) 19th March      (3) 17th to 19th, March      (4) anytime
15. Where will the book fair take place?
   (1) a bookstore  (2)* Cultural Center  
   (3) a school  (4) parking lot

16. What will educational speakers be doing at the Book Fair?
   (1) *giving talks  (2)selling ticket  
   (3)reading books  (4)signing names

17. Which of the following will be at the book fair?
   (1) There will be only children’s books, no adult books.  
   (2) No authors will be present at the Fair.  
   (3) Exhibition of children of all ages  
   (4) *Demonstration of latest book designs.

18. The Book Fair’s opening hours on Saturday are:
   (1) 9:30-5:00  
   (2)* 9:30-4:00  
   (3) between 9:30 and 5:00  
   (4) all day

19. Who will pay $250 for a ticket to the fair?
   (1) an author  (2) an educator  
   (3) an adult  (4)* a student

20. How can you buy tickets?
   (1) *calling 03-8786789  
   (2) calling 03-03-8782222  
   (3) go in person  
   (4) calling the hotline 03-8782222 to reserve an all-day parking space.

* This item was not used with the middle school student version of the test.
Our test design possesses the following two features. Firstly, it conforms to what McNamara (1996) described as a “weak performance test” in that getting the knowledge-based accurate answer is not as important as making the appropriate language choice in a discourse. A weak performance test therefore, tests the ability of an individual to communicate in a general sense without referring to any specific knowledge area. A “strong performance test”, in contrast, would be a test that success would be judged in terms of both the subject area knowledge and the English language proficiency.

Secondly, our design catches the essence of what Lado (1961) proposed in the design of language tests. Lado thought language tests should concentrate on testing “control of the problems”, which refer to the units and patterns of the target language that do not have a counterpart in the student’s native language, “…….” (Lado, 1961, p.24). Our test items were designed in a way that there are no direct counterparts for the words and phrases in Chinese as they appear in the items, but the expressions are so commonly used in English language that if one fails to understand what they mean in a discourse, one can not reasonably be perceived as functional in the English language.

The instrument is designed in a way that there will be no controversy in scoring for all native English speakers as well as for people with a functional level of knowledge in using English as a foreign language. Given the authors’ familiarity with the English-speaking conditions in Southeast Asia, had such a test been given to students of middle school level in Hongkong, Malaysia or Singapore, it would most likely yield percent-corrects greater than .7 for all items.

Subjects

We sought and acquired cooperation from six English teachers (two from Year 7, Year 8, and Year 9, respectively) in two large middle schools (one in Taipei and one in Kaoshiung) for the purposes of the current study. The instrument was given by the first author following all the procedures of a standardized test administration. A total of 229 test papers were collected, of which 79 are from Year 7, 78 from Year 8, and 72 from Year 9. A couple months later the same instrument (with one additional item added to it) was given to students enrolled in undergraduate statistics, educational measurement, and psychology classes in a teachers college following the same procedure as earlier. The ages of students ranged from 18 to 21. A total of 112 test
papers were collected.

There is roughly equal representation of both genders across all years in middle schools, whereas the male-to-female ratio in teachers college is roughly 1 to 2. It is expected that the students of the teachers college would outperform the middle school students due to the nature of the admission process of selecting students into the higher education institution. All students were given 30 minutes time to complete the test, but for all middle school classes, most students (over 80%) completed in less than 20 minutes, whereas for teachers college classes, most students completed in less than 15 minutes.

Statistical Analyses

Two types of statistics are analyzed concerning the psychometric properties of an item: percent-correct and item discrimination. The former is measured by the percentage of examinees answering the item correctly, also commonly referred to as the difficulty level of the item. The higher the percent correct is, the easier the item is. The latter is measured by point-biserial correlation coefficient which gives indication of the strength of association between the item score and the total scale score. Discrimination in the context of psychometric properties means the ability of an item to separate high and low ability students and is thus a desirable trait. The higher the point-biserial correlation, the more discriminating the item is. In our current study, the percent correct and discrimination statistics for all twenty items are given in Table 1, which provides the empirical basis for all subsequent statistical analyses.

A split-plot ANOVA design is used to analyze the effect of work sample (Type) and the effect of level of schooling (Level) with Type serving as the whole-plot treatment variable and Level serving as the sub-plot variable. The two dependent variables here are percent correct scores and discrimination indices. It is noteworthy to mention that the units of analyses in our design are the items of the test, not the students taking the test. Type is a whole-plot variable because the items in three work samples are different items. Each item is used four times, which correspond to measurements taken at Year 7, 8, 9 and college time points sequentially. This design is commonly referred to as the repeated measures design in the context of social science research methodology (Shaughnessy, Zechmeister & Zechmeister, 2006). The advantage of using such a design lies in its ability to test both the main effects (Type and Level) and their perceivable interaction effect simultaneously.
Table 1.

Item analysis for three work samples by four levels of schooling

<table>
<thead>
<tr>
<th>Item No</th>
<th>Percent Correct</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 7</td>
<td>Year 8</td>
</tr>
<tr>
<td>1_1</td>
<td>0.203</td>
<td>0.167</td>
</tr>
<tr>
<td>1_2</td>
<td>0.392</td>
<td>0.308</td>
</tr>
<tr>
<td>1_3</td>
<td>0.342</td>
<td>0.513</td>
</tr>
<tr>
<td>1_4</td>
<td>0.519</td>
<td>0.577</td>
</tr>
<tr>
<td>1_5</td>
<td>0.190</td>
<td>0.321</td>
</tr>
<tr>
<td>2_1</td>
<td>0.266</td>
<td>0.449</td>
</tr>
<tr>
<td>2_2</td>
<td>0.228</td>
<td>0.423</td>
</tr>
<tr>
<td>2_3</td>
<td>0.304</td>
<td>0.410</td>
</tr>
<tr>
<td>2_4</td>
<td>0.354</td>
<td>0.449</td>
</tr>
<tr>
<td>2_5</td>
<td>0.152</td>
<td>0.192</td>
</tr>
<tr>
<td>2_6</td>
<td>0.418</td>
<td>0.538</td>
</tr>
<tr>
<td>2_7</td>
<td>0.430</td>
<td>0.436</td>
</tr>
<tr>
<td>2_8</td>
<td>0.228</td>
<td>0.308</td>
</tr>
<tr>
<td>3_1</td>
<td>0.468</td>
<td>0.474</td>
</tr>
<tr>
<td>3_2</td>
<td>0.392</td>
<td>0.397</td>
</tr>
</tbody>
</table>
### Results

**General English Ability**

Despite the lack of representativeness of our sample, it is disappointing to see that Taiwanese students’ English communicative ability is poor, as evidenced by an average of 6.651 on a 0-19 scale for the middle school students, barely higher than a third of the total possible score. Given that these results were based upon students chosen from the two highest social economic status areas in the country, the general condition of English communicative ability may be much worse for the rest of the country. The results from the college sample are no more comforting either, with an average of 13.661 on a 0-20 scale. Despite the fact their mean score is substantially higher than that of the middle schools’ students, we must also take into account their age difference (they are on average 4 years older than the middle school students) and they are considered as top of the cream in their age group when compared to the rest of the population their age. If their performance is seen as shown, the rest of the population would certainly perform much worse.

For both middle school and college samples, the total raw scores are skewed slightly in opposite directions (skewness indices of .704 and -.622 for middle schools and teachers college, respectively). The opposite directions of skewness were expected because the test is considered difficult for the middle schools students, and is relatively easy for the college students. The internal consistency based reliability indices as measured by Cronbach’s alpha coefficient are .810 and .672, respectively. This pattern of opposite skewness is observed across all three work samples. In terms of item percent correct, the averages of middle school samples ranged from .32 to .37,
whereas it was .69 for the college sample. With respect to item discrimination, the
averages of middle school samples ranged from .54 to .59, whereas it was .50 for the
college sample. It is noteworthy that while our items are considered difficult for the
middle school students, they nevertheless do possess reasonable discriminatory power
when it comes to assessing English communicative ability.

Item Percent Correct Analysis

The results of ANOVA for item percent correct are reported in Table 2. It is not
surprising that Level was found to be such a highly significant effect because students
are expected to improve as they receive more education. Despite the non-significant
indication for the main effect of Type, its interpretation is complicated by the presence
of the significant interaction effect between Type and Level (F=4.16, P<.002). We
shall take a more detailed look at the nature of this interaction effect by examining the
plot of means for three work samples across four levels of schooling as shown in
Figure 2.

As seen, the interaction effect may mainly be attributed to the differential role
semantic competence had played across four levels of schooling in comparison to the
pragmatic competence. The percent correct associated with semantic items is roughly
the same or lower than the pragmatic items for middle school students (Year 7 to Year
9). This situation turned around in which semantic items become easier as compared
to their pragmatic counterparts.
Table 2.

ANOVA results for percent correct analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>2</td>
<td>0.006</td>
<td>0.003</td>
<td>0.04</td>
<td>0.961</td>
</tr>
<tr>
<td>Item(Type)</td>
<td>16</td>
<td>1.175</td>
<td>0.073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>3</td>
<td>1.677</td>
<td>0.559</td>
<td>78.92</td>
<td>0.000</td>
</tr>
<tr>
<td>Type*Level</td>
<td>6</td>
<td>0.177</td>
<td>0.029</td>
<td>4.16</td>
<td>0.002</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>0.340</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Item(Type) is used as the error term for testing the effect of Type, whereas Error is used to test the effects of Level and Type*Level.

Figure 2.
Item Discrimination Analysis

The results of ANOVA for item discrimination are reported in Table 3. No interaction effect was found. Both Type and Level main effects are significant, with Type as a highly significant main effect (F=9.89, P<.002) and Level as a moderately significant effect (F=2.80, P<.05). The plot of means for three work samples across four levels of schooling is shown in Figure 3.

As seen, Semantic items are more discriminating for the middle school students. This result is certainly not surprising because middle school students are more used to items measuring semantic competence than they are to the items measuring pragmatic competence. After all, the so called high English ability students at the middle school level are commonly perceived to mean those who can memorize answers to fixed patterns of tests in which the contents have all been previously exposed to them (all words must have been taught in class). For the college students, items measuring semantic competence seem to be equally discriminating as the pragmatic items. This is also expected because students tend to increase their semantic competence as they progress from middle school to college. At this stage, students with higher semantic competence will also tend to have higher pragmatic competence.

Table 3.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>2</td>
<td>0.993</td>
<td>0.497</td>
<td>9.89</td>
<td>0.002</td>
</tr>
<tr>
<td>Item(Type)</td>
<td>16</td>
<td>0.804</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>3</td>
<td>0.085</td>
<td>0.028</td>
<td>2.80</td>
<td>0.050</td>
</tr>
<tr>
<td>Type*Level</td>
<td>6</td>
<td>0.104</td>
<td>0.017</td>
<td>1.71</td>
<td>0.139</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>0.484</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Item(Type) is used as the error term for testing the effect of
Type, whereas Error is used to test the effects of Level and Type*Level.
Figure 3.

Conclusions

Our research hypotheses are shown to be supported. Three conclusions can be reached based on the results of the current study:

1. Items measuring both semantic and pragmatic competence will become easier as students progress from middle school to college. This is a positive interpretation of our empirical findings.
2. Judging from percent correct statistics while taking into account the age factor, Taiwanese students English ability in terms of both semantic and pragmatic competence is far from being satisfactory.
3. With respect to item discrimination, semantic items are more discriminating than pragmatic items. This phenomenon is undesirable because it implicates the current English assessment fails to take into account students pragmatic competence in the current notion of English assessment.
Discussion

We hope we have clearly articulated the need for attention to the pragmatic aspect of Taiwanese students’ English ability, particularly in the current atmosphere of near-neurotic accountability that is required of the English teachers (or teachers at all levels in general). We tried to be direct, straightforward as well as cogent in pointing out that English teachers should keep in mind that language itself is never static or dehumanizing. Neither should the assessment of English ability be. We realize that the current English testing and assessment in Taiwanese schools reflects the overall cultures of power-wielding and top-down authority of the past generation, and it takes time for the young generation of teachers to break out of this conformity. We simply wish to allude to the perception of powerlessness by the parents under the current system in the hope of improving their children’s English ability, and also the sociocultural consequences of failing to meet the demand for better pragmatic ability in the use of the English language.

Many people blame the uses and misuses of objective types of tests for the low English ability of Taiwanese students. In fact, objective tests in and of themselves are not the cause of the current conditions. The culprit is, rather, that most English teachers are inflexible with the contents of their tests. In our observation, English tests are often used as a disciplinary tool, forcing students to produce uniform answers; while at other times they become the tool of school policy in the schemes of accountability, compliance, and standardization. We have demonstrated in this paper that objective testing, when designed properly, can be employed to assess the students’ English ability from the perspective of pragmatic competence. We call on the Taiwanese English teachers to come to the recognition that English, as a language, should reflect diversity growing in the current English speaking world. English education in this country should broaden its scope in order to generate critical thinking and creative individuals who celebrate the diversity and change of the English language. It is our sincere hope that the day will come that English teachers in this country will turn tests into a means for change as well as devices for promoting agendas that reflect the need of our people. Will Taiwanese teachers change their traditional practices in the assessment of English ability to embrace the challenges of the future? Will Taiwanese students’ English ability catch up with those of our counterparts in Hongkong and Malaysia? Hopefully yes we believe, but not yet.
References