INSTRUCTORS' PERCEPTIONS OF THE CONTENT VALIDITY OF THE ENGLISH LANGUAGE EXAMS AT NIĞDE UNIVERSITY

A THESIS PRESENTED BY

MAHMUT METIN AKSAN

TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES IN

PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

BILKENT UNIVERSITY

JULY, 2001

ABSTRACT

Title: Teachers' Perceptions of the Content Validity of the English Language Exams at Niğde University

Author: Mahmut Metin Aksan

Thesis Chairperson: Dr. Hossein Nassaji
Bilkent University, MA TEFL Program

Committee Members: Dr. William E. Snyder
Dr. James C. Stalker
Bilkent University, MA TEFL Program

The purpose of this study is to find out the teachers' perceptions of whether the content of the tests items reflect the content of the coursebook and their teaching. An equally important purpose is to find out whether the instructors follow the coursebook content in their teaching.

This study attempted to find out these research questions:

1. Are there any significant differences among Niğde University English Instructors in terms of whether they think final test items reflect the content of the coursebook?
2. Are there any significant differences among Niğde University English Instructors in terms of whether they think final test items reflect their teaching?
3. What is the relationship between teachers' perceptions of the relationship of the tests to the content of the coursebook and their teaching?

Data was collected from sixteen English teachers of Niğde University through a questionnaire which was consist of 40 test items chosen randomly among five final English tests which were taken from English instructors.

To analyse the results of the questionnaire, quantitative analysis methods were used in this study. Chi-square statistical analysis was used to analyse the data.

The results of the first research question indicate that instructors generally think the test items reflect the content of the coursebook they use. In other words the instructors feel the tests have content validity. Results of the second research question show that instructors generally think that the test items reflect their teaching. The results of the third research question indicate that instructors follow the coursebook content in their teaching.

A further examination of the test items indicates that there are 13 problematic items within the tests. The categories of these problematic items are : multiple correct answers, response cues, no correct answer, number of options, and translation.

There are some precautions that can be taken in order to reduce the number of problematic items in the tests. One of them is peer review which instructors check each other's tests. An in-service training may be very useful for the instructors on testing and it may be a good solution of raising the content validity of the tests.

BILKENT UNIVERSITY

INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES

MA THESIS EXAMINATION RESULT FORM

JULY 13, 2001

The examining committee appointed by the Institute of Economics and Social Sciences

for the thesis examination of the MA TEFL student

Mahmut Metin Aksan

has read the thesis of the student.

The committee has decided that the thesis of the student is satisfactory.

Thesis Title	: Instructors' Perceptions of the Content Validity of the English Language Exams at Niğde University
Thesis Advisor	: Dr. William E. Snyder Bilkent University, MA TEFL Program
Committee Members:	: Dr. James C. Stalker Bilkent University, MA TEFL Program
	Dr. Hossein Nassaji Bilkent University, MA TEFL Program

We certify that we have read this thesis and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Masters of Arts.

Dr. Hossein Nassaji
(Chair)

Dr. William Snyder
(Committee member)

Dr. James Stalker
(Committee member)

Approved for the
Institute of Economics and Social Sciences

Kürşat Aydoğan
Director
Institute of Economics and Social Sciences

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Dr. William Snyder, without whose invaluable guidance this thesis would never come true.

I am deeply grateful to Dr. James C. Stalker and Dr. Hossein Nassaji who provided me support and encouragement through out this research and this program.

I would also like to thank to my classmates, without them I would never be able to survive.

My greatest and special thanks go to my wife and my daughters. Thank you for your patience.

Behind every successful man, there is a strong woman (I have three).

To Semiha, Dilara and Destina

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1 INTRODUCTION

Introduction

The purpose of this study is to find out the perceptions of English instructors as to whether English language final tests at Niğde University reflect the content of the coursebook and instructors' teaching, and whether there is a relation between coursebook content and teaching.

The validity of tests is one of the most important issues in preparing tests. "Validity is the degree to which a test actually tests what it is intended to test" (Kitao and Kitao, 1996, see also Carroll & Hall, 1985, Harris, 1969, Hughes, 1989, Weir 1988). Brown (1996) claims that we have to prepare valid tests to improve our teaching because teaching and testing are related to each other. We expect valid tests increase positive washback (Messick, 1996). Washback is the effect of testing on teaching (Hughes, 1989).

One aspect of validity is content validity. Content validity is the degree of the representativeness of course content in test (Brown, 1996, Carroll & Hall, 1985, Harris, 1969, Hughes, 1989, Weir 1988). The content of the test must be the same or similar with the content of the course for the classroom tests. Heaton (1990) talks about how the importance of classroom tests. "Classroom tests are the most important tests for teachers because the reason of classroom tests is to find out how well the students have mastered the language areas and skills which have just been taught" (p. 9). Heaton adds that content validity is important in classroom tests because our students expect that questions related to the course content will be asked. Their success is bound to the tests. If they encounter questions that are not from the course content, they will not be able to answer, to be successful. Because of the difficulty of the tests, students may not want to learn English or because of the difficulty of the tests teachers may want to overload the students. Their attitudes towards English also might change in negative way, and they will never want to learn English again.

Kitao and Kitao (n.d.) suggest that experts should be in charge of validating the content of the tests. Experts may be teachers themselves (Brown, 1996), so it is important to teach to the teachers how to evaluate a test or how to prepare a valid test. Also colleagues, directors of the institutions may check the tests in terms of the content validity. Experts opinions are based on the criteria which they used in deciding whether a test content valid or not. These criteria may come from literature or from teachers' meetings. Teachers are the only experts at schools for deciding the validity of their individual tests or their colleagues tests, because there is not any separate department for the determination of the content validity of the tests. Colleagues know what they teach during the course more than other people.

However, instructors were considered as experts in content validating the English language tests at Niğde University since the test content validation was the responsibility of the instructors in the institution.

Context of the Study

I started teaching English at Niğde University in 1998. Since I started to work at this university I have been listening to my colleagues, teachers from other departments and my students. Generally their thoughts about English lessons, learning English, teaching English and the usefulness of English are similar to each other. Many of the teachers, students and administrators think that the way of teaching English at Niğde University is not right. They also think that students are not learning English, that their time in the classroom is wasted while learning English. In classroom discussions about English it is easy to see that many of the students do not like English because of their background. Some of them did not learn English in high school. Some of them had bad experience with English and English teachers. Some of them think that English is not their language and it is nonsense to learn it. Some of the teachers of other departments tell students that they will not need English in their future so they do not need to study English very much. After listening to these kinds of statements from their department teachers, students don't show any eagerness to learn English in the classrooms.

At the beginning of the year we have an exemption from English lessons test in all of the university. Some of the students pass this exam. After this test we assume all of the students who could not pass the test are at the elementary level. As a result of this we teach English at the beginning level. However, this is only an assumption. Since some of the good students who are good at English did not take the exemption test, they attend the classes. So we have students at different levels in our classes although we have an exemption test at the beginning of the year.

Based on their teaching, English instructors at Niğde University prepare their midterm or final tests individually. Some of the teachers use questions related to the extra materials they use in the class. Some of the teachers ask questions only from the teachers' book of the textbook. According to the students, some tests are difficult and some tests are easy. Some of the students pass their classes because their teachers ask very easy questions, but some of them can't pass their classes because their teachers ask very difficult questions in the tests.

Statement of the Problem

This study is a descriptive study looking at instructors' the perceptions of English tests at Niğde University in terms of reflecting the content of the coursebook and instructors' teaching. As I mentioned in the previous section, content validation requires experts' opinions (Kitao & Kitao, n. d.), but, at Niğde University, every teacher prepares his or her own tests. Every teacher follows his or her own way in preparing these tests. Some of them ask multiple-choice questions, some of them ask open-ended questions, others ask reading comprehension questions. So all of them are in charge of validation of their tests individually in terms of content.

Significance of the Study

This study will contribute to the literature on teachers' perceptions of testing. This study is intended to be useful for the English teachers of Niğde University. Not considering the content validity of English tests of Niğde University causes many doubts among the students, teachers and administrators. This study will provide evidence about the perceptions of the content validity of English tests at Niğde University. If there is a problem it will be easier to solve it because being aware of the sources of the problem is a place to start. The results of this may also give incentive to instructors at Niğde University to improve the quality of the teststhey prepare.

Research Questions

This study will address the following research questions:

1- Are there any significant differences among Niğde University English Instructors in terms of whether they think final test items reflect the content of the course book?

2-Are there any significant differences among Niğde University English Instructors in terms of whether they think final test items reflect their teaching?

3- What is the relationship between teachers' perceptions of the relationship of the tests to the content of the coursebook and their teaching?

CHAPTER 2 LITERATURE REVIEW

Introduction

The main purpose of this study is to find out whether there are differences among the Niğde University English language instructors' perceptions of the final tests of 2000-2001 academic year in terms of reflecting the content of the course book and their teaching. These tests were prepared individually by the English instructors of Niğde University. Since this study is about the content validity of the tests at Niğde University, the aim of this literature review is to provide information about content validity and validity as well as testing and teaching. So most of the literature reviewed in this chapter is about the purposes of testing, impact, washback, validity, content validity, and test items.

Purposes of Testing

In a second or foreign language learning classroom, teachers should know what their students achieve in terms of learning. Also, learners want to see some record of their performance and their development. For both these reasons, assessment is very important. The performance of the students in a course can be learned in different ways. Giving them projects, oral examining and testing are some of these ways. One of the ways of assessment is testing and testing gives teachers, administrators and students a real thing to keep in their hands. These real things are generally the grades that are taken from the tests (Bachman, 1991, Brown, 1996, Henning, 1987, Hughes, 1989). These grades show teachers and administrators how much their students achieved and which subjects students did not learn. Satisfactory knowledge of students' performances is very important for the teachers and the administrators, because course planning is generally based on this performance

(Brown, 1996). He gives other purposes of testing such as teachers deciding to make changes in their syllabuses to be sure that students learn everything in their courses. Also, the good and bad sides of the methodology applied in the schools can be seen easily after having this knowledge in hands. Test scores are very important because teachers decide which of the students pass their classes, students can learn their achievement, parents have information about their students' achievement, and administrators evaluate syllabi and the curriculum depending on these test scores (Brown, 1996).

Tests results give teachers, administrators, parents and students important feedback on whether the students are achieving or not. So, as teachers, we have to be very careful while preparing tests. We should try to make our students to be ready for the tests (Hughes, 1989). For these reasons validity and especially the content validity of classroom achievement tests are very important. But before going into validity and content validity we should take a look at the impact and washback of tests.

Impact and Washback

Impact is the effect of tests on society and individuals. Test taking and use of test scores have two kinds of impact. One is macro and the other one is micro. Macro impact is the effect on society, education system and micro impact is the effect on individuals (Bachman & Palmer, 1996). Bachman and Palmer give three aspects of how testing procedures affect test takers.

"1. the experience of taking and, in some cases, of preparing for the test,

2. the feedback they receive about their performance on the test, and

3. the decisions that may be made about them on the basis of their test scores" (p.31).

Washback is an aspect of impact according to Bachman & Palmer (p.30). Hughes gave another definition of washback. It is the effect of testing on teaching (Hughes, 1989). Generally the main purpose of teaching at schools is final evaluation of students. This final evaluation includes the subjects in the syllabi or the coursebooks. Teachers ask questions about the important pedagogical items according to them. This also makes teachers to teach the pedagogical items which are asked in the tests. Tests are very important for the students and also for the teachers, because teachers decide whether their students pass their classes or not depending on the results of these tests. These results have a great effect on students' future.

Validity

In order to take right decisions about the students depending on tests, tests have to have validity. Valid tests measure what they intend to measure (Carroll & Hall, 1985, Harris, 1969, Henning, 1987, Hughes, 1989, Weir, 1988). In other words, when you want to test a specific skill or knowledge you have to ask questions related to that skill or knowledge. Nobody can say that his or her tests are valid without the control of experts (Brown, 1996, Kitao & Kitao, n.d.). These experts can be teachers who have had training in testing, colleagues who know the subject of the test or perhaps administrators who have experience in testing (Brown, 1996). While preparing tests we can ask our colleagues to check the items of the tests in terms of validity, because they also teach the same things with us, such as the same book. Administrators may also be good experts if they also have the knowledge of English (Brown, 1996). Experts have to prepare criteria while checking the test items in terms of validity (Henning, 1987). These criteria must cover the content of the course. They can take the syllabus or the curriculum and compare with the items. The items should be checked carefully.

"Assessment of teacher practice must be both valid and reliable if it is to be believed and trusted. Validity relates to the question of whether or not one assesses what one claims to or intends to assess" (Wenning, 2000). According to Wenning (2000) validity is a must in tests, and teachers' practice must be built around validity and reliability. Before teaching something to the teachers about education, educators of teachers should teach how the teachers can make their tests reliable and valid.

There are different kinds of validity that we should be careful of while preparing our tests. These are the construct validity, face validity, and content validity of the tests (Carroll & Hall, 1985, Henning, 1987, Harris, 1969, Hughes, 1989, Weir, 1988). In this research I am interested in the content validity of the tests in Niğde University. So I will talk only about content validity.

Content Validity

If a test's content is a good representative of the course content in terms of language skills and structures then we can say these tests have content validity (Hughes, 1989, Henning, 1987, Kerlinger, 1973). If you want to test some grammar points such as past perfect tense, you can't ask present perfect tense in this test and claim that it shows knowledge of past perfect tense. Also, in a listening test can not ask a reading item to the students. A listening test without listening questions doesn't have the content validity. Content valid tests refer to the tests which test the content of the course. The tests which are content valid measure the things taught during the course (Brown, 1996).

When we want to test a part of a course or to learn the overall achievement of a student in a course we have to prepare a content valid test (Harris, 1969). The items of the tests which are prepared for a specific course must have the same pedagogical content as the course content, not a different one. Besides this we can not leave out some subjects which we taught during the course. Suppose that we taught 50 items during a course but we asked only 30 items in the test. This test doesn't have the content validity because it is not the representative of the course content (Henning, 1987; Kerlinger, 1973). But on the other hand Innes and Straker (2000) think in another way. They simply say that one test may have content validity without having all the content of the course. But the degree of the content validity may decrease. So, to increase our test's content validity degree, we should include as much as subjects we taught during the course in the order of frequency in the course. Both of the definitions may be possible. But the important thing is where we are applying these views. For the mid-term tests we are not supposed to include all the material we taught during the course. But for the final tests it is better to include all the materials we taught during the course, because final tests are the overall measurement of the students.

The content selection of the tests is very important, because this selection carries the content validity. This process is easier in preparing achievement tests than preparing proficiency tests, because in the achievement test preparation process there is a course instruction and course content. But in proficiency test preparing process there may be the length of testing as a guiding constraint (Henning, 1987).

Experts are important while validating the tests. They also should know the content of the course to be able to validate the tests in terms of content.

> In order to investigate content validity, testers must decide whether the test is a representative sample of the content of whatever the test was designed to measure. To address this issue, testers or some of their colleagues usually end up making some sort of judgements. This content validation process may take many forms, depending on the particular language teaching situation and staff, but the goal should always be establish an argument that the test is a representative sample of the content that the test claims to measure (Brown, 1996).

For that reason the most talented are teachers because teachers know what they taught and how they taught in the course. A closer examination of the tests by the teachers will increase the content validity of the tests (Alderson, Clapham & Wall, 1995, Brown, 1996, Henning, 1987).

Studies on Content Validity

There have been some studies on determining content validity of tests. One of them was done by Scott, Stansfield and Kenyon (1996) on the listening summary translation exam (LSTE)- Spanish version administered by the Federal Bureau of Investigation (FBI). The purpose of this study was to gather quantitative evidence of the reliability and validity of LSTE. The subjects of this study were 67 examinees. Both forms of LSTE-Spanish were given in one sitting at each of seven FBI field offices in the USA and Puerto Rico. In addition to the LSTE-Spanish a self-assessment questionnaire on which each examinee was asked to estimate his or her ability to perform summary translation tasks. The comparison of the results of LSTE-Spanish and the self-assessment questionnaire were similar to each other. The results of this study are the evidence for LSTE's content validity. In the case of the LSTE-Spanish, evidence for its content validity is found in the tasks examinees are asked to perform to demonstrate their ability in listening summary.

Another study about use of test method characteristics in the content analysis and design of EFL proficiency tests by Bachman, Davidson and Melanotic (1996). The research reported on the use of content analysis in the comparision of two different EFL proficiency test batteries that was conducted as part of the Cambridge-TOEFL Comparability Study (CTCS) of the comparison of multiple forms of a single EFL proficiency test battery. The purpose of the research was to describe the content of multiple forms of the First Certificate in English. It investigated these tests' content comparability and relationship between test content and item statistics. The other purpose of this study was to provide feedback to the tester. The test maker may use this feedback in the revision of test specifications. For this study, Cambridge First Certificate of English Test, Paper 1 (Reading Comprehension) was used. For the content analysis five raters were include in this study. Raters consistency was examined using two different methods: 1) variance components from generalizability study with raters, and characteristics nested within raters were estimated for all forms combined and 2) rater agreement proportion (RAP) was used as a second approach. When the both of the methods are taken together, the results indicate a very high level of rater agreement in validating the content of CTCS.

Another study was done Teasdale (1996). It conducted in the course of the development of an English language test for newly-qualifying Air Traffic Control trainees. Air Traffic Control language has its own content. For that reason the tests have to have the content of this language area. In order to define this area of language use and specify the domain a Needs analysis of the work-specific language use of Air Traffic Controllers was conducted. The domain specification was carried out through transcription of recordings of authentic Air Traffic Control speeches and through a questionnaire. There were 76 responses from 15 different countries. A fixed category questionnaire with open-ended slots for comments was designed to investigate the test characteristics. The results of this study indicate that the contents of the test do not totally reflect the language needs in communication in Air Traffic Control field.

Another study was done by Harun Serpil at Anadolu University about the content validity of the midterm achievement tests. The purpose of this study was to analyze the content validity of the first semester midterm tests. To investigate the teachers' perceptions of the tests' representation of the classroom material content, their opinions were elicited by using a questionnaire. To learn the objectives of the courses, coordinators responsible for each course interviewed. The results were conflicting. The instructors of the listening and grammar courses thought their the tests reflect their course content. But the analysis of course objectives showed that they were not specific enough and their overall agreement with the tests' content was low.

The studies above about the validity and the content validity of the tests. The last study which was done by Harun Serpil is closely related to my study. In fact there are not so many studies done on the content validity I hope my study will contribute to this area of testing.

Test Items

A test item is the smallest part of a test (Brown, 1996, p. 49). Gathering the items in one place develops a test. So, discovering the content validity of a test it is very important to analyze the items in the test, one by one. There are some possible problems with test items which we should avoid them in order to make our tests more valid and reliable. We can place problematic items under the categories that Henning (1987) and Brown (1996) discussed in their books. According to Henning (1987), there are several possible kinds of errors that are made in tests. These are mixed response, response cue, number of options, nonsense distracters, review options, trick questions, common knowledge response, matching material, redundancy, and medium of response.

Mixed Response

If an item intends to measure a specific part of grammar such as simple past tense, but it has more than one possible choice with simple past tense among the options this means that it doesn't have content validity (Henning, 1987). Henning gives an example and its solution for this kind of problematic items (p. 43-44).

Example: John ………….. flowers to the party last night.

a) Carries c) lifts

b) Carried d) lifted

The response options have to be set in this way :

a) carries c) is carrying

b) carried d) has carried

Response Cues

It is very difficult to avoid response cues in preparing test item distractors. This means that students can choose the right answer among the options without using real knowledge of the item being tested. "Students who have had much prior exposure to these kinds of examinations may be said to have developed test 'wiseness: that is, such students may be capable of selecting the correct opinion independently of any knowledge of the content field being tested" (Henning, 1987, p.

43). because they may have developed a test wiseness as a result of being exposed too much to the same kind and style tests. Henning (1987) talks about three different kinds of response cues. These are length cues, convergence cues and inconsistent distractor cues. Length cues may provide evidence for students to think that the longest option is the correct one. Convergence cues are when different categories of distractors converge to provide students with a basis for making a choice (in semantic or phonological) of the correct form. An inconsistent distractor may make the students to think that the very different option among the distractors is the wrong one (Henning, 1987).

Number of Options

Too many or too few choices in an item cause validity and reliability problems. Consider true/false questions. There are only two options for the students and one of them is the right answer. With great possibility students choose the right answer. Consider a listening test with five choices. It is impossible to follow the listening material and trying to find the right answer among the choices (Henning, 1987).

Nonsense Distracters

Nonsense options have two problems with them. First of all "…they tend to be weak distracters" (Henning, 1987, p. 45), and secondly "…they have negative 'washback' on instruction" (Henning, 1987, p. 46). The purpose of tests is to get information about students' achievement. It is not their duty to teach something during the test, especially wrong things.

Review Options

Options, which require review to the other options, are not good options because they make students lose time while referring back to the other options. For example: The stranger had left his native land because he

a) wished to seek his fortune.

b) wanted to avoid his creditors.

c) preferred the new land.

d) none of the above

e) *a* and *b* but not *c* above

f) *b* and *c* but not *a* above (Henning, 1987, p. 46).

Trick Questions

Trick questions cause invalid measurement and bad pedagogy. This questions are asked in the tests because teachers want to show their cleverness and to ensure test difficulty (Henning, 1987). For example:

When is not appropriate not to be absent from class?

a) when you are sick

b) when you are young

c) while class is in session

d) whenever the teacher is angry (Henning, 1987, p. 46).

For this example, the use of the double negative makes it difficult to understand the question.

Redundancy

In order to gain time in the test, test makers should avoid repetitions in the response options (Henning, 1987). For example:

Students should study harder,

a) because they should pass their classes

b) because they should learn everything

c) because they should be a good person for their countries

Instead of the question like the one above we can ask,

- Students should study harder because they should

a) pass their classes

b) learn everything

c) be a good person for their countries

Brown (1996) also talks about the similar categories. He also discussed about item format analysis. "In item format analysis, testers focus on the degree to which each item is properly written so that it measures all and only the desired content" (Brown, 1996, p. 50). He also gives a checklist questions for doing item format analysis.

Checklist Questions	Yes	No
1- Is the item format correctly matched to the purpose and content of the item?	….	….
2- Is there only one correct answer?	….	….
3- Is the item written at the students' level of proficiency?	….	….
4- Have ambiguous terms and statements been avoided?	….	….
5- Have negatives and double negatives been avoided?	….	….
6- Does the item avoided giving clues that could be used in answering other items?	….	….
7- Are all parts of the item on the same page?	….	….
8- Is only relevant information presented?	….	….
9- Have race, gender, and nationality bias been avoided?	….	….
10- Has at least one other colleague looked over the items?	….	….

Figure 1 Item Format Analysis Checklist from Brown (1996, p. 51)

Conclusion

The main aim of this literature review was to provide information about the purposes of testing, impact, washback, validity, content validity, and test items.

Content validity is a primary concern about which teachers need to be careful while preparing tests. Creating content validity requires a lot of time. The tests which are prepared at Niğde University individually by the teachers do not have so much time to be checked. So, validity, especially content validity, should be examined in the tests given at Niğde University.

CHAPTER 3: METHODOLOGY

Introduction

This study examined the perceptions of Niğde University English teachers about the 2000-2001 academic year first term final tests prepared individually by the English teachers of Niğde University in terms of reflecting the content of the course book and their teaching. Reflecting the content of the course book is the basis to determine the content validity in this study. Reflecting their teaching was asked to understand if there is a difference between the course book content and their teaching.

Content validity in tests is one of the most important issues about which teachers have to be concerned. To find out the perceptions of Niğde University English teachers about the content validity of the English tests of Niğde University, a questionnaire was prepared for the teachers, asking their thoughts about the test items in terms of reflecting the content of the course book and their teaching.

Participants

This study was conducted in Niğde University Foreign Languages Department. The participants in this study were the English instructors at Niğde University Foreign languages department.

There were twenty English instructors at Niğde University Foreign Languages Department, but the questionnaires were administered to only the sixteen English instructors who were currently teaching. Of these sixteen participants, there were three female instructors and thirteen male instructors. Only one of these instructors has a master's degree. Another instructor is currently taking a Masters course, which he will finish next year. The others have undergraduate degrees in ELT.

Their length of experience varies from one year to twenty years. Five of the instructors are new instructors who began to teach last academic year. The age range of the participants of this study is from 26 to 45. There are six participants who previously had taken a language-testing course among these sixteen participants.

Materials

In this part the procedure for choosing the tests, test items, and designing the research questionnaire are explained.

The items of the questionnaire were chosen randomly from five different final tests from the first term of 2000-2001 academic year. There were sixteen different final tests given at Niğde University because there were sixteen English teachers and they prepare their own tests individually. I asked all teachers to give me their final tests for use in this study. However I received only five final tests, so these were used for the study. If I had received more than five final tests I would have used all of those received.

I had in total 129 test items. I decided to choose a sample of items from each of the tests. Eight items were chosen randomly from each of the tests, so that there were forty test items in the questionnaire. I took 25% of the total items of Test 1, 40% of the total items of Test 2, 22% of the total items of Test 3, 40% of the total items of Test 4, and 25 % of the total items of Test 5.

After choosing the items from the tests, a questionnaire was prepared for the English teachers at Niğde University (See Appendix A). The test items used in the questionnaire were mixed in order to ensure that those five teachers who gave their tests wouldn't recognise their test items when they saw them in the questionnaire.

For each test item two different questions were asked:

a) How well does this test item reflect the content of the textbook you use?

b) How well does this test item reflect the content of your teaching?

A Likert-scale was used for the answers of these questions. The choices were arranged in the order, "not at all", "badly", "somewhat well", "well", and "very well". I asked the two questions for each of the test items because I wanted to learn if the teachers' perceptions were different regarding the relationship between the tests and content of the course book and their teaching.

The current students of MA TEFL program checked the format of the questionnaire. I didn't pilot the questionnaire with the teachers at Niğde because the number of participants was small.

Procedure

Data were collected in approximately one week in April, 2001. There is a stuff room for the Foreign Languages Department in the Science of Economics Faculty which the instructors of Niğde University Foreign Languages Department use as an office. The teachers of Foreign Languages Department of Niğde University go there only on their on-duty days. All of them have classes at different faculties in different places in Niğde. They don't need to be at this office every day. I waited for them at this office. When they came to the office after their courses or for their office hours, I gave them the questionnaires. This process took five days.

The female teachers were very eager to answer the questions in the questionnaire. However, it was very difficult to get the male teachers to answer the questionnaires. Sometimes I had to supervise them while they completed the questionnaire. The questionnaires were administered and collected on the same day. The teachers were not allowed to take them home.

The teachers whom I took the example tests from were also included in the study, because, firstly, all the test questions were randomly mixed to be sure that they would not recognise their own questions in the questionnaire. Secondly, they couldn't be left out because of the small number of participants. If they had been excluded there would have been only 11 participants for this study.

The reason I had the English teachers of Niğde University rate the test items was that they all had taught the same book and all of them knew what was in the book. I thought that they could easily evaluate the items. Also it was impossible to find experts at Niğde University to rate the items of the questionnaire. In this study I wanted to know the perceptions of Niğde University English teachers on the content validity of the tests prepared by themselves. For that reason it was a must to have them rate the items.

Data Analysis

In this study the main instrument for the data collection was a questionnaire. For analysis, the test items which were used in the questionnaire were regrouped again according to the tests they were taken from. After this procedure, responses were counted according to the test items and frequencies were recorded. Tables were prepared for each of the tests. The tables were made for each the tests and for both questions a and b. Chi-square statistical analysis were used to find out whether there are differences among the perceptions of the English instructors of Niğde University about the final tests of 2000-2001 academic year first term. These tables will be shown and discussed in Chapter 4.

CHAPTER 4 DATA ANALYSIS

Introduction

This study was done to find out if there are differences among the perceptions of Niğde University English instructors about the tests prepared individually by English instructors of Niğde University in terms of reflecting the content of the coursebook and their teaching. The study was done through a questionnaire prepared for English instructors of Niğde University. The questionnaire included 40 test items, which were chosen randomly from among five tests which were taken from the English instructors of Niğde University.

The data analysis presented in this chapter consists of three sections. The first section presents the views of the instructors on the content validity of the tests and the relationship between their teaching and the tests. The second section reports the perceptions of the instructors on the relationship between the individual test items and the content of the coursebook and also their teaching. The third section gives the analysis of the individual items which were found to be problematic after the analysis of the test items in the second section of this chapter.

Chi-square analysis was used to find out whether there are differences among the responses to the questionnaire items by Niğde University English instructors. In the next section, the results of the analysis of the instructors' perceptions of the tests will be shown.

Perceptions of the Tests

Table 1 and table 2 show the perceptions of the English instructors of Niğde University about the relation of the tests which were prepared in 2000-2001 academic year to the content of the coursebook and their teaching.

Table 1.

The Perceptions of Instructors of the Final Tests in Terms of Reflecting the Content of the Coursebook

	Not at all	Badly	Somewhat well	Well	Very well	Total
Test 1	10	28	35	31	24	128
Test 2	15	33	36	31	13	128
Test 3	7	27	32	47	15	128
Test 4	13	31	28	32	24	128
Test 5	10	34	29	38	16	127
Total	55	153	160	179	92	639

Note. df = 16, Chi-square = 17.75

The chi-square analysis of Table 1 was not significant, showing that there were no differences among the instructors in the evaluation of the tests. However, a closer examination of the responses suggests that instructors think that the tests reflect the coursebook. The sum of the responses under ‘very well’ and ‘well’ (n = 271) is greater than the sum of the responses under ‘not at all’ and ‘ badly’ (n = 208). In addition, 160 responses indicated that instructors felt that the tests reflect the coursebook at least somewhat well. Overall, these results show that the tests reflect the content of the coursebook according to the instructors’ responses.

Table 2

The Perceptions of the Instructors of the Final Tests in Terms of Reflecting Their Teaching

	Not at all	Badly	Somewhat well	Well	Very well	Total
Test 1	11	27	33	33	24	128
Test 2	9	36	29	34	20	128
Test 3	6	26	33	39	24	128
Test 4	13	33	27	30	25	128
Test 5	8	29	33	36	21	127
Total	47	151	155	172	114	639

Note. df = 16, Chi-square = 8.37

The chi-square analysis of Table 2 was not significant, showing that there were no differences among the instructors in the evaluation of the tests. However, a closer examination of the responses suggests that instructors think that the tests reflect their teaching. The sum of the responses under ‘very well’ and ‘well’ (n = 286) is greater than the sum of the responses under ‘not at all’ and ‘ badly’ (n = 198). Here, as well, there were 155 responses showing that the tests reflect their teaching at least somewhat well.

The results of Table 1 indicate that instructors generally feel the tests are content valid according to the definition used in this study. In other words the tests reflect the coursebook. The results of Table 2 show that instructors generally feel the tests reflect their teaching. The correspondence of the sum of the responses of the two tables shows that instructors appear to follow the coursebook in their teaching.

Even though the results above are generally positive, there are substantial negative responses that we can not ignore. In order to try to understand these negative responses, I will look at the individual items within tests to learn the instructors’ perceptions of the items. The results of this analysis will be shown in the next section.

Instructors’ Perceptions of the Test Items

Tables 3 and 4 show the perceptions of the English instructors of Niğde University about the relation of the items of Test 1 which was prepared individually in 2000-2001 academic year to the content of the coursebook and their teaching.

Questions from the tests were arranged randomly for the questionnaire and have been regrouped here for analysis. The numbers of the questions in the tables show the order of the questions in the questionnaire.

Table 3

The Perceptions of Instructors of the Items of Test 1 in Terms of Reflecting the Content of the Coursebook

	Not at all	Badly	Somewhat well	Well	Very well	Total
Question 3	0	2	5	4	5	16
Question 7	0	0	5	9	2	16
Question 10	1	2	4	5	4	16
Question 22	1	2	7	2	4	16
Question 24	1	2	5	4	4	16
Question 27	1	5	4	3	3	16
Question 31	1	7	3	4	1	16
Question 40	5	8	2	0	1	16
Total	10	28	35	31	24	128

Note. df = 28, Chi-Square = 51.05, p < .01

The chi-square analysis of Table 3 is significant at .01 level. This indicates that there were differences among the instructors in their assessment of the test items. If we look deeply into the responses of the instructors, they suggest that, while the instructors are generally positive about the items in this test, there may be problems with some of test items in terms of reflecting the content of the coursebook. For my purposes, a test item will be considered problematic, if the sum of the responses under ‘not at all’ and ‘badly’ is greater than the sum of the responses under ‘well’ and ‘very well’. In Table 3, the items, which have a greater number of the responses under ‘not at all’ and ‘badly’, are 31 and 40. These items will be discussed in the section ‘Problematic Items’.

Table 4

The Perceptions of the Instructors of the Items of Test 1 in Terms of Reflecting Content of Their Teaching .

	Not At All	Badly	Somewhat Well	Well	Very Well	Total
Question 3	0	1	5	5	5	16
Question 7	0	0	6	6	4	16
Question 10	1	3	2	6	4	16
Question 22	1	2	7	3	3	16
Question 24	1	2	5	5	3	16
Question 27	1	6	3	4	2	16
Question 31	1	6	3	4	2	16
Question 40	6	7	2	0	1	16
Total	11	27	33	33	24	128

Note. df = 28, Chi-square = 49.54, p < .01

The chi-square analysis of Table 4 is significant at .01 level, showing that there were differences among the instructors in their evaluation of the test items. When we examine the responses they suggest results similar to those in Table 3, including that there may be problems with some of the test items in terms of reflecting the instructors’ teaching. These problematic items, according to the definition I gave before, are items 27, 31 and 40. These items will be discussed in the section ‘Problematic Items’.

Tables 5 and 6 show the perceptions of the English instructors of Niğde University about the relation of the items from Test 2 to the content of the coursebook and their teaching.

Table 5

The Perceptions of Instructors of the Items of Test 2 in Terms of Reflecting the Content of the Coursebook

	Not At All	Badly	Somewhat Well	Well	Very Well	Total
Question 1	2	6	4	4	0	16
Question 15	0	0	5	6	5	16
Question 20	1	2	6	6	1	16
Question 26	0	3	4	5	4	16
Question 29	1	8	4	2	1	16
Question 32	2	5	5	3	1	16
Question 34	2	5	5	4	0	16
Question 37	7	4	3	1	1	16
Total	15	33	36	31	13	128

Note. df = 28, Chi-Square = 50.92, p < .01

The chi-square analysis of Table 5 is significant at .01 level. This shows that there were differences among the instructors in their evaluation of the test items. If we compare the sum of the responses under ‘not at all’ and ‘badly’ (n = 48) with the sum of the responses under ‘very well’ and ‘well’ (n = 44) we can see that instructors are more negative about how well the items in this test reflect the coursebook. This is reflected in the fact that more than half the items are problematic according to the definition given earlier. The problematic items of this test are 1, 29, 32, 34, and 37. These items will be discussed later.

Table 6

The Perceptions of the Instructors of the Items of Test 2 in Terms of Reflecting Content of Their Teaching.

	Not At All	Badly	Somewhat Well	Well	Very Well	Total
Question 1	1	6	3	5	1	16
Question 15	0	0	3	7	6	16
Question 20	0	2	6	5	3	16
Question 26	0	3	5	4	4	16
Question 29	0	9	3	2	2	16
Question 32	1	6	3	4	2	16
Question 34	1	6	4	5	0	16
Question 37	6	4	2	2	2	16
Total	9	36	29	34	20	128

Note. df = 28, Chi-Square = 55.57, p < .01

The chi-square analysis of Table 6 is significant at .01 level. This supports that there were differences among the instructors in their evaluation of the test items. In this case, comparing the sums of the responses under 'not at all' and 'badly' (n = 45) with 'very well' and 'well' (n = 54), we can see that the instructors are more positive about how well the items of this test reflect their teaching. However, there are still problems with some of test items, according to the definition I gave before. These problematic items are again 1, 29, 32, 34, and 37.

In Test 2, the number of the problematic items is more than the half of the total number of the selected items from Test 2. There are five problematic items out of eight selected items. This indicates that this test has a great amount of problems with it, which we cannot ignore.

Tables 7 and 8 show the perceptions of the English instructors of Niğde University about the relation of the items from test 3 to the content of the coursebook and their teaching.

Table 7

The Perceptions of Instructors of the Items of Test 3 in Terms of Reflecting the Content of the Coursebook

	Not At All	Badly	Somewhat Well	Well	Very Well	Total
Question 2	2	4	3	7	0	16
Question 6	0	3	5	8	0	16
Question 9	0	2	4	7	3	16
Question 17	1	2	2	7	4	16
Question 19	1	2	4	7	2	16
Question 21	1	3	5	4	3	16
Question 25	1	3	4	5	3	16
Question 39	1	8	5	2	0	16
Total	7	27	32	47	15	128

Note. df = 28, Chi-Square = 28.52

The chi-square analysis of Table 7 is not significant. This shows that there were no differences among the instructors in their assessment of the test items. This suggests that the instructors generally viewed the test items in the same way in terms of reflecting the content of the book they taught during the academic year. However, according to my definition of the problematic items the responses show there is still one problematic item, item 39, in Test 3. It will be discussed in the section 'Problematic Items'.

Table 8

The Perceptions of the Instructors of the Items of Test 3 in Terms of Reflecting Content of Their Teaching.

	Not At All	Badly	Somewhat Well	Well	Very Well	Total
Question 2	1	5	3	5	2	16
Question 6	2	3	3	6	2	16
Question 9	0	2	3	7	4	16
Question 17	0	2	3	7	4	16
Question 19	0	2	4	8	2	16
Question 21	1	2	6	4	3	16
Question 25	1	3	4	2	6	16
Question 39	0	8	7	0	1	16
Total	5	27	33	39	24	128

Note. df = 28, Chi-Square = 37.03

The chi-square analysis of Table 8 is not significant. This shows that there were no differences among the instructors in their evaluation of the test items. This suggests that the instructors generally viewed the test items in the same way in terms of reflecting their teaching. Only item 39 has a greater number as the sum of the responses under 'not at all' and 'badly' than under 'very well' and 'well'. So this item will be discussed in the section 'Problematic Items'.

Tables 9 and 10 show the perceptions of the English instructors of Niğde University about the relation of the items from Test 4 to the content of the coursebook and their teaching.

Table 9

The Perceptions of Instructors of the Items of Test 4 in Terms of Reflecting the Content of the Coursebook

	Not at all	Badly	Somewhat well	Well	Very well	Total
Question 4	5	4	1	3	3	16
Question 11	1	5	6	3	1	16
Question 13	1	6	1	3	5	16
Question 16	0	5	2	6	3	16
Question 18	1	3	6	3	3	16
Question 23	0	0	5	6	5	16
Question 36	1	2	5	5	3	16
Question 38	4	6	2	3	1	16
Total	13	31	28	32	24	128

Note. df = 28, Chi-Square = 39.85

The chi-square analysis of Table 9 is not significant. This shows that there were no differences among the instructors in their evaluation of the test items. This suggests that the instructors generally viewed the test items in the same way in terms of reflecting the content of the coursebook they taught in 2000-2001 academic year. However, there are still two items which have a greater number of responses under ‘not at all’ and ‘badly’ than ‘very well’ and ‘well’. These are items 4 and 38, which will be discussed later.

Table 10

The Perceptions of the Instructors of the Items of Test 4 in Terms of Reflecting Content of Their Teaching .

	Not at all	Badly	Somewhat well	Well	Very well	Total
Question 4	3	7	2	1	3	16
Question 11	1	5	5	4	1	16
Question 13	2	5	1	4	4	16
Question 16	0	5	5	3	3	16
Question 18	0	4	4	5	3	16
Question 23	0	0	5	5	6	16
Question 36	2	2	4	5	3	16
Question 38	5	5	1	3	2	16
Total	13	33	27	30	25	128

Note. df = 28, Chi-Square = 36.27

The chi-square analysis of Table 10 is not significant. This shows that there was no difference among the instructors in their evaluation of the test items. This suggests that the instructors generally viewed the test items in the same way in terms of reflecting the content of their teaching. However, there are still two items, which have a greater number of responses under 'not at all' and 'badly' than 'very well' and 'well'. These items are again numbers 4 and 38. These items will be discussed later.

Tables 11 and 12 show the perceptions of the English instructors of Niğde University about the relation of the items from test 5 to the content of the coursebook and their teaching.

Table 11

The Perceptions of Instructors of the Items of Test 5 in Terms of Reflecting the Content of the Coursebook

	Not at all	Badly	Somewhat well	Well	Very well	Total
Question 5	1	4	3	5	2	15
Question 8	0	4	2	7	3	16
Question 12	0	4	4	5	3	16
Question 14	0	2	2	8	4	16
Question 28	0	6	2	5	3	16
Question 30	2	8	2	3	1	16
Question 33	2	1	9	4	0	16
Question 35	5	5	5	1	0	16
Total	10	34	29	38	16	127

Note. df = 28, Chi-Square = 51.23, $p < .01$

The chi-square analysis of Table 11 is significant at .01 level. This shows that there were differences among the instructors in their evaluation of the test items. This suggests that while the instructors are generally positive about the items in this test, there may be problems with some of test items in terms of reflecting the content of the coursebook they used during the academic year. According to my definition there are two problematic items in this test. These are items 30 and 35. These items will be discussed in the following section.

Table 12

The Perceptions of the Instructors of the Items of Test 5 in Terms of Reflecting Content of Their Teaching.

	Not at all	Badly	Somewhat well	Well	Very well	Total
Question 5	0	3	3	7	2	15
Question 8	1	2	4	5	4	16
Question 12	0	3	4	5	4	16
Question 14	0	0	5	7	4	16
Question 28	0	6	2	4	4	16
Question 30	2	8	3	1	2	16
Question 33	1	1	8	5	1	16
Question 35	4	6	4	2	0	16
Total	8	29	33	36	21	127

Note. df = 28, Chi-Square = 48.08, $p < .05$

The chi-square analysis of Table 12 is significant at .05 level. This shows that there were differences among the instructors in their evaluation of the test items. This suggests that while the instructors are generally positive about the items in this test, there may be problems with some of test items in terms of reflecting the content of their teaching. According to the definition I gave before there are two problematic items in Test 5. These items are again 30 and 35. These items will be discussed in the section ‘Problematic Items’.

Problematic Items

The analysis of the items in the previous section found 13 problematic items out of the 40 total. In Chapter 2, I looked at the kinds of problems that there can be in an item, based in part on Henning’s (1987) and Brown’s (1996) discussions. I will group the problematic items from the tests which I used in the questionnaire by category. The categories that I will use are: multiple correct answers, response cues, no correct answer, number of options, and translation.

Multiple Correct Answers

Teachers and students may think in different ways while evaluating an item. A student may think that one option is correct but the teacher may decide that it is incorrect (Brown, 1996, p. 50). In the items I analysed in the previous section there are some problematic items that have more than one correct answer among their choices. In this situation students may not choose the answer that the teacher wants. These items are 27, 29, 31, 37, and 38.

Item 27: Boşluklara "some-any-a-an-this-these-that-those" kelimelerini uygun şekilde yazın: (Translation: Write the appropriate words in the blanks "some-any-a-an-this-these-that-those")

Is there .... garden?

In this item there are two possible answers "any" and "a" among the options.

Item 29: You ....wear comfortable clothing.

a) shouldn't b) don't have to c) should

In this item there are three options and grammatically all of them can be correct answer.

Item 31: Boşluklara "some-any-a-an-this-these-that-those" kelimelerini uygun şekilde yazın. (Translation: Write the appropriate words in the blanks "some-any-a-an-this-these-that-those)

Is .... your brother over there?

In item 31, the expected correct answer is 'that' but students may think that the correct answer is 'this'. Both of them are grammatically correct.

Item: 38: ....do you travel to work?

a) what b) when c) where

This item is similar to item 29. It also has three options and two possible correct answers, “b” and “c”, among the options. The instructor did not give enough explanation about the context of this item.

For all of the items above, the test maker did not give any contexts that make students understand or choose the expected option. One more sentence might make the context clear.

While the items above have more than one correct answer among the options, item 37 has the same answer twice among its options and will be discussed below in the section Number of Options. While the instructor should be very careful while preparing options for a test item, it may be good to have a colleague check the items in the test before administering them. Then, these kinds of problems may be eliminated. Also, by using Brown’s (1996) checklist, the quality of the items in tests may be increased.

Response Cues

Henning (1987) talks of how the different shapes of a test item may provide ‘response cues’ that point to correct answer. When the students have had the same kinds of tests during the term or year they may have become familiar with the test types and options. This familiarity causes students to evaluate the options without having to use knowledge of the language.

> "Students who have had much prior exposure to these kinds of examinations may be said to have developed 'test wiseness'; that is such students may be capable of selecting the correct option independently of any knowledge of the content field being tested" (Henning, 1987, p. 44).

This test wiseness may help students to choose the correct answer among the choices without having any knowledge about the subject asked in the test.

Item 39 is the only test item in the questionnaire which students can understand the correct answer among the options in this way.

Item 39: Claude didn't..................... in Canada

a) livedb) use to live c) used to live d) used to living

Item 39 has four options and only one of the options was not written in Simple Past Tense. While teaching Simple Past we tell students that if we use 'did' or 'did not' after the subject of the sentence we do not need to use the past form of the verb. Option 'b' has no past tense form of verb, but the other three options have past tense form of verbs. If we look at the item we can see that 'did not' is used with the subject. This is a good cue for a careful student who has developed test wiseness.

No Correct Answer

Neither Henning nor Brown talk about items having no correct answer in their books. This is because they might have thought that instructors have to put the correct answer among the options. After analysing the test items I found one item which has no correct answer among the options. This item is item 4.

Item 4: Frank lives in Leeds. He lives ....two other boys are students.

a) in b) at c) to

In item 4, there is no correct answer. Perhaps the instructor was not careful while preparing this item. To avoid these kinds of mistakes in our test items we can show them to a colleague to check the items in our tests. A careful examination may solve the problem.

Number of Options

"Care should be taken to ensure the proper of options for any given set of items. Problems may arise when the number of response options are too few or too many" (Henning, 1987, p. 45). In other words if we put too few or too many options for multiple-choice items we decrease the validity of our tests. A listening test item with more than five options may cause a lack of attention to the listening material. Also, true/false items may give students a chance to choose the right answer among the choices in 50% chance.

Henning does not give any exact number for the proper amount of numbers. Normally, in Turkey, teachers and instructors prepare test items with four or five options. Teachers' negative responses generally for the following items (1, 4, 29, 32, 34, 37 and 38) may be because they have only three distracters. In addition, they rated these items as problematic items because they have other problems which were discussed under the different categories in this section.

Item 1: What .... you ....if you ....?

a) will/do/fail b) are/doing/fail c) do/do/fail

Item 4: Frank lives in Leeds. He lives ....two other boys are students.

a) in b) at c) to

Item 29: You ....wear comfortable clothing.

a) shouldn't b) don't have to c) should

Item 32: I...............you when lunch ....ready.

a) will/is b)call/come c) 'll call/is

Item 34: Paul plays guitar and sings.........................

a) only b) especially c) too

Item 37: wealthy:.........................

a) poor b) poor c) generous

Item 37 has two possible correct answers. In fact, these two correct answers are the same answers. This means that students have only two options to choose from. In other words, they have a 50 % chance of choosing the correct one. A careful examination of the items may solve this problem. A double check by the tester may also be a good solution in solving the problem.

Item: 38: ....do you travel to work?

a) what b) when c) where

There are also items which have too many options. These items are 27 and31.

Item 27: Boşluklara “some-any-a-an-this-these-that-those” kelimelerini uygun şekilde yazın: (Translation: Write the appropriate words in the blanks “some-any-a-an-this-these-that-those”)

Is there .... garden?

Item 31: Boşluklara “some-any-a-an-this-these-that-those” kelimelerini uygun şekilde yazın. (Translation: Write the appropriate words in the blanks “some-any-a-an-this-these-that-those)

Is .... your brother over there?

Above two items have too many options with them. This might be a problem for the instructors to rate them as problematic items.

Translation

There is one item which asks the students to translate a given Turkish word into English. This is a one-word translation.

Item 40: Kelimerin İngilizce karşılıklarını yazın: (Translation: Write the English equivalent of the given words)

Arasında (Among or between) ....

If instructors want to ask any translation questions it might be more useful and appropriate to ask full sentences. Instructors at Niğde University might not like one-word translations items.

Conclusion

After examining the results of the study I can simply say that the tests generally reflect the content of the coursebook and the content of teaching applied in the classrooms according to the English instructors of Niğde University. The book content and teaching also appear to be closely related to each other. In other words instructors appear to follow the course book in their teaching. Still, a closer analysis of the data revealed that 13 of 40 items were problematic in some ways. The amount of these problematic items cannot be ignored while validating the tests.

I will discuss the results, implications and the limitations of this study in the next chapter.

CHAPTER 5 CONCLUSION

Overview of the Study

This study discovered the perceptions of Niğde University English teachers of the first term final tests of 2000-2001 Academic Year at Niğde University. Sixteen English teachers of Niğde University participated in this study. Data was collected using a questionnaire administered among the participants. This study attempted to find out the English instructors' perceptions of English language tests at Niğde University, in terms of reflecting the content of the coursebook and instructors' teaching and whether there is a relation between the coursebook content and their teaching.

Summary of the Findings

In this chapter I discuss the results of my data by answering each of my research questions. The first research question is:

Are there any significant differences among Niğde University English Instructors in terms of whether they think final test items reflect the content of the course book?

To find the answer of this question "How well does this item reflect the content of the course book you use?" was asked for each of the test items in the questionnaire. In Table 1, sum of the responses under 'well' and 'very well' (n = 271) is greater than the sum of the responses under 'not at all' and 'badly' (n = 208). This shows that most of the teachers are positive about the tests in terms of reflecting the content of the coursebook. As I defined the content validity of the tests as the teachers feel that the degree of representativeness of the coursebook content, the results show that tests are content valid.

My second research question is:

Are there any significant differences among Niğde University English Instructors in terms of whether they think midterm test items reflect their teaching?

To find the answer of this research question "How well does this item reflect the content of your teaching" was asked for each of the item in the questionnaire. In Table 2, sum of the responses under 'well' and 'very well' (n = 286) is greater than the sum of the responses under 'not at all' and 'badly' (n = 198). This shows that in general the thoughts of the English instructors at Niğde University about the tests are positive in terms of reflecting their teaching. This indicates that the tests reflect the teaching going on in classrooms.

My third research question is:

What is the relationship between teachers' perceptions of the relationship of the tests to the content of the coursebook and their teaching?

To find the answer of this question the responses which were given to the "How well does this item reflect the content of coursebook" were compared with the responses which were given to the "How well does this item reflect your teaching". This comparison indicates that instructors appear to follow the coursebook content in their teaching. Although there were some doubts that the instructors did strictly follow the coursebook content, the findings of this study revealed that the instructors do appear to follow the coursebook content.

Discussion

This study was started because of perceived problems at Niğde University. These problems included doubts about teaching and testing. These doubts about the teaching and the testing caused many problems including uncertainty about the quality of instruction among instructors and students.

This study attempted to investigate the problems about teaching at Niğde University through instructors' perceptions of the tests. The results of this research show that there are not any problems with the tests and instructors' teaching. However, deeper analysis of individual items revealed potential problems. As shown in chapter 4, there were 13 (33%) problematic items out of 40 sample items. All of the problematic items in this questionnaire were discussed in chapter 4. These problematic items suggest possible problems in test design.

Implications

Pedagogical Implications

Peer review is one of the most helpful procedures in preparing valid tests. Teachers may show their tests to their colleagues to be sure that there are not any errors in their tests. Many of the teachers hesitate to show their tests to their colleagues, because they don't want their errors and mistakes to be known by their colleagues. But the instructors should be in cooperation with each other in order to make teaching and testing more useful to their students.

The analysis of the test items individually show that there were some problematic items (see Chapter 4, Problematic Items) which are the evidence of the problems with test design. Test design requires training. The instructors who had training on testing might have rated the items differently from the instructors who did not have any training. But we do not know in what way they rated the items. It may be useful to organise an in-service training for teachers on testing with specialists in this field.

Implications for Further Study

This study covered only the final tests of the first term of the 2000-2001 at Niğde University. It will be very useful to do a research on a whole year's tests prepared by the teachers of Niğde University. This study also looked at only the instructors' perspectives. For future study, it may bring different results if students' perceptions of tests are considered as well. Students may rate the test items differently, more positively or more negatively than instructors.

In this study the background of the participants was not considered. The age wasn't considered also. Having master or doctoral degree might have played a great role. So, for future study, considering the background of the participants may be very useful for the results of the study.

Limitations of the Study

This study discovered only the instructors' perceptions of the 2000-2001 academic year first term final tests at Niğde University, in terms of reflecting the content of the book and their teaching. This limits the generalizability of the study. In addition, small number of the participants makes the study less generalizable.

Some of the teachers were not so eager in answering the questionnaire questions. Sometimes I had to supervise them. Some of them might have rated in positive ways because they did not want to judge their colleagues. This was also a limitation for this study.

REFERENCES

Alderson, C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. New York: Cambridge University Press.

Bachman, L. F., Davidson, F., & Milanovic, M. (1996). The use of test method characteristics in the content analysis and design of EFL proficiency tests. *Language Testing, 13*, 125-149.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Hall Regents.

Carroll, B. J., & Hall, P. J. (1985). *Make your own language test*. New York: Pergamon Press.

Carey, L. M. (1988). *Measuring and evaluating school learning*. Boston, MA: Allyn & Bacon.

Davies, A. (1990). *Language test validation* Oxford: Basil Blackwell Ltd.

Fulcher, G. (1997). An English placement test: Issues in reliability and validity. *Language Testing, 2,* 113-139

Harris, P. D. (1969). *Testing English as a second language*. New York: McGraw Hill Inc.

Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House Publishers.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press

Innes, E., & Straker, L. (1998). Validity and reliability. [Online] Available: http://curtin.edu.au/curtin/dept/physio/pt/staff/straker/publications/1999Work5Validity.html

Kitao, S. K., & Kitao, K. (1996). Validity and reliability [Online] Available: http://ilc2.doshisha.ac.jp/users/kkitao/library/article/test/design.htm#validity

Scott, M. L., Stansfield, C. W., & Kenyon, D. M. (1996). Examining validity in a performance test: The listening summary translation exam (LSTE)-Spanish version. *Language Testing, 13,* 83-109

Teasdale, A. (1996). Content validity in tests for well-defined LSP domains: an approach to defining what is to be tested in, M. Milanovic, & N. Saville (Eds.), *Studies in Language Testing 3* (pp. 211-230). Cambridge: Press Syndicate of the University Cambridge

Wenning, C. J. (2000). Validity and reliability in performance assessment. [Online] Available: http://phy.ilstu.edu/ptefiles.html

APPENDIX

Dear Participant,

I am a graduate student at Bilkent University MA TEFL Program in Ankara. I am working on my thesis and this questionnaire is for data collection of my thesis. You are expected to answer all the questions without leaving blank any of them. Your answers will be used only for this research and will not be announced. Thank you for your cooperation.

Mahmut Metin AKSAN

A) BACKGROUND INFORMATION

1- Sex :............... Male Female

2- Age:.................... 21-25 26-30 31-35 36-40 41-45 46-50

3- Have you ever had a course on testing? Yes No

B) **TEST TEMS**

The section below contains 40 sample test items. Below each item are two questions asking your opinion on a five-point scale. Please circle the response that best reflects your opinion.

1- What….. you……. if you ……..?

a) will/do/fail b) are/doing/fail c) do/do/will fail

- How well does this item reflect the content of the text book you use?

  Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

  Not at all badly somewhat well well very well

2- The driver _________a speeding ticket. The police are right behind him.

a) gets b) is getting c) is going to get d) will get

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

3-Where were you born? ....

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

4-Frank lives in Leeds. He lives …….. two other boys who are students.

a)in b) at c) to

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

5- ……………… are you?

I'm 14.

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

6-What______________ these days?

a) are you doing b) do you do c) you are doing d) you do

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

7- Aşağıdaki soruları cevaplandırın

How old are you? ....

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

8- ………………does this dres cost?

$65.00

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

9- When you get to the corner, ___________left.

a) is turning b) turn c) turning d) turns

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

10- What time is it?

10:55 ....

- How well does this item reflect the content of the text book you use?

Not at all   badly   somewhat well   well   very well

- How well does this item reflect the content of your teaching?

Not at all   badly   somewhat well   well   very well

11- They cook a meal for their friends and they go out …….. the pub

a)for b) after c) by

- How well does this item reflect the content of the text book you use?

Not at all   badly   somewhat well   well   very well

- How well does this item reflect the content of your teaching?

Not at all   badly   somewhat well   well   very well

12- …………….. did Eve go to Italy?

Last month

- How well does this item reflect the content of the text book you use?

Not at all   badly   somewhat well   well   very well

- How well does this item reflect the content of your teaching?

Not at all   badly   somewhat well   well   very well

13- Nigde …… in Türkiye

a) is b) amn't c) isn't

- How well does this item reflect the content of the text book you use?

Not at all   badly   somewhat well   well   very well

- How well does this item reflect the content of your teaching?

Not at all   badly   somewhat well   well   very well

14- Angela regularly (take) ………………………. the bus to work.

- How well does this item reflect the content of the text book you use?

  Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

  Not at all badly somewhat well well very well

15- I enjoy ………… in the sea very much.

a) to swim b) swimming c) swim

- How well does this item reflect the content of the text book you use?

  Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

  Not at all badly somewhat well well very well

16- I ………on holiday. I'm at work

a) am b) amn't c) am not

- How well does this item reflect the content of the text book you use?

  Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

  Not at all badly somewhat well well very well

17- Will you be home tomorrow night? No, _____________

a) I don't b) I'm not c) I will d) I won't

- How well does this item reflect the content of the text book you use?

  Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

  Not at all badly somewhat well well very well

18- My teachers ……….. very funny.

a) is b) are c) are'nt

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

19- I'll see you_______________

a) at the moment b) in an hour c) last night d) usually

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

20- John managed …………his room before his mother came home.

a) tidy b) to tidy c) tidying

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

21- One day last March, I ____________ a very starnge letter

a) did get b) got c) used to get d) was getting

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

22- Boşluklara “in-on-at” kelimelerini yazın.

……….4 p.m.

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

23-There aren’t………… good restaurants in our town.

a) some b) any c) an

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

24-Boşluklara “in-on-at” kelimelerini yazın.

…………weekends.

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

25-Who ____________ yesteday at the store?

a) did you see b) did you use to see c) you saw d) you were seeing

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

26- I'd like ……….to India.

a) to go b) go c) going

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

27-Boşluklara "some-any-a-an-this-these-that-those" kelimelerini uygun şekilde yazın.

Is there …………garden?

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

28-Penny and Tom never (have) ……………………….meat for dinner.

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

29-You ……………….wear comfortable clothing.

a) shouldn't b) don't have to c) should

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

30-Aşağıdaki çizili bölümlere göre cümleleri soruya çeviriniz.

The twins chose <u>chocolate ice-cream</u> ………………….

- How well does this item reflect the content of the text book you use?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

- How well does this item reflect the content of your teaching?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

31- Boşluklara "some-any-a-an-this-these-that-those" kelimelerini uygun şekilde yazın.

Is ……………. your brother over there?

- How well does this item reflect the content of the text book you use?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

- How well does this item reflect the content of your teaching?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

32- I………you when lunch …………ready.

a) will/is b) call/come c) 'll call/ is

- How well does this item reflect the content of the text book you use?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

- How well does this item reflect the content of your teaching?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

33- While John (walk) ……………to school yesterday, he (meet) ……….. Judy.

- How well does this item reflect the content of the text book you use?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

- How well does this item reflect the content of your teaching?

<u>Not at all</u> <u>badly</u> <u>somewhat well</u> <u>well</u> <u>very well</u>

34- Paul plays the guitar and sings ……………

a) only b) especially c) too

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

35-Aşağıdaki boşlukları going to kullanarak verilen fillerle tamamlayınız.

look rain fail eat get

That man …………………wet, because he hasn't got an umbrella.

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

36- There's……….. newsagent's opposite the post office.

a) some b) any c) a

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

37- wealthy = ………..

a) poor b) poor c) generous

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

38- ………….. do you travel to work?

a) what b) where c) when

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

39- Claude didn't ______________ in Canada

a) lived b) use to live c) used to live d) used to living

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well

40- Kelimelerin İngilizce karşılıklarını yazın.

Arasında ………....

- How well does this item reflect the content of the text book you use?

Not at all badly somewhat well well very well

- How well does this item reflect the content of your teaching?

Not at all badly somewhat well well very well