2006-07-26

Paraprofessionals in the Classroom: Creating a Test Bank

Linda Hansen
Brigham Young University - Provo

PARAPROFESSIONALS IN THE CLASSROOM:

CREATING A TEST BANK

By

Linda L. Hansen

A thesis submitted to the faculty of

Brigham Young University

In partial fulfillment of the requirements for the degree of

Master of Science

Department of Counseling Psychology and Special Education

Brigham Young University

August 2006

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Linda L. Hansen

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

_____	_____
Date	Betty Y. Ashbaker, Chair
_____	_____
Date	Mary Anne Prater
_____	_____
Date	Lynn K. Wilder

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Linda L. Hansen in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____ _____

Date

Betty Y. Ashbaker
Chair, Graduate Committee

Accepted for the Department

_____

Tina T. Dyches
Graduate Coordinator

Accepted for the College

_____

K. Richard Young or Barbara Culatta
Dean, McKay School of Education
Associate Dean, McKay School of Education

ABSTRACT

This purpose of this thesis was to create a valid, reliable, fair test bank for the textbook *Paraprofessionals in the Classroom* (Ashbaker & Morgan, 2006). This textbook was written specifically for the education of paraprofessionals. Further education beyond high school is one of the options recommended by the No Child Left Behind Act of 2001 to assist paraprofessionals in becoming highly qualified to work with children with special needs. Extensive research was conducted by the researcher on what constituted a good test bank. Excerpts were selected from seven chapters of the textbook. Questions and a rating scale for each question were written for each excerpt. The mentors selected paraprofessionals from various fields of special education to answer and rate the questions. Each paraprofessional read the excerpts and questions from three different chapters and then rated each question for difficulty level. The answers and question ratings were analyzed, the questions revised as necessary, and the test bank was completed.

ACKNOWLEDGEMENTS

My thanks to Dr. Betty Ashbaker who helped me to open the castle door and started me down the halls of learning. My thanks to my husband who was at each corner to cheer me on. My thanks to my children who knew I could find my way and had such faith in my ability. My thanks to my instructors who cheerfully invited me into their rooms to share with me their life's treasures of knowledge. My thanks to my beloved students with whom I could share my nuggets of wisdom learned and thus help them to start on their own journeys. My thanks to my precious grandchildren who are the jewels in the throne room and who help me to realize that each child has such great potential and, since Heavenly Father never gives up on anyone, neither will I.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

CHAPTER ONE

INTRODUCTION

History

The reauthorization of the Individuals with Disabilities Education Act of 1997 (IDEA) allows appropriately trained paraprofessionals to work with students under the direction of a teacher. The reauthorization of IDEA acknowledges that the educational system may use "paraprofessionals and assistants who are appropriately trained and supervised …to be used to assist in the provision of special education and related services to children with disabilities" (20 U.S.C. §1412 (a)(15)(B)(iii).

Historically, some of the first paraprofessionals who were hired by the State of Michigan were college educated (Pickett, 1999). Their task was to do paperwork, not work with the students directly. In the ensuing years, paraprofessionals have not been required to have a formal education. Their informal education has been on the job training or training done by the district for which they worked. With the No Child Left Behind Act of 2001, which will hereafter be referred to as NCLB, the roles and responsibilities of the paraprofessional had come under scrutiny as legislators questioned whether these individuals were truly qualified to work with this student population. Giangreco, Yuan, McKenzie, Cameron, and Fialka (2005), researchers of special education, state that "the least qualified staff members are teaching students with the most complex learning characteristics" (p. 29). Regardless of the severity of the student's needs, there were no formal standards for paraprofessionals' qualifications.

Congress concluded that a national standard was needed so that the quality of paraprofessional service would be appropriate for the students served. Recent

reauthorization of NCLB now requires training or an associate's degree or equivalent for Title I paraprofessionals. Since many paraprofessionals were already employed by the time NCLB was enacted and had years of experience, some states have given them the opportunity to submit a portfolio documenting the training they have received during their years of service. In this chapter, a definition of a paraprofessional will be provided with an explanation their roles and responsibilities and how this affects the purpose of this study.

Definition of a Paraprofessional

Para means "alongside of" (Pickett, 1999). The special education and general education teachers, related servers, parents, and Local Education Agency (LEA) comprise the Individualized Education Program (IEP) team and it is the team's responsibility to design an IEP for each student with special needs. This consists of reviewing the student's present levels of academic performance and writing goals to improve this performance. Paraprofessionals work alongside the team members, delivering this program. Another definition given by NCLB defines a paraprofessional and required duties as:

For the purposes of Title I, Part A, a paraprofessional is an employee of an LEA (Local Education Agency) who provides instructional support in a program supported with U.S. Department of Education (DOE), NCLB, (2001), Part A funds.

> 'Paraprofessionals who provide instructional support' include those who (a) provide one-on-one tutoring if such tutoring is scheduled at a time when a student would not otherwise receive instruction from a teacher; (b) assist with classroom management, such as by organizing instructional materials; (c) provide

> instructional assistance in a computer laboratory; (d) conduct parental involvement activities; (e) provide instruction support in a library or media center; (f) act as a translator; or (g) provide instructional support services under the direct supervision of a highly qualified teacher. (DOE, Title I, Section 1119(g)(2)).

Paraprofessionals are important members of the education team who work under the direction of a teacher, implementing the program created by the IEP team for the benefit of the individual students with special needs. By fulfilling their roles and responsibilities, they can assist the student with special needs to progress on the goals written in the IEP.

Roles and Responsibilities of a Paraprofessional

Paraprofessional duties include ". . . under the supervision of teachers and other professional practitioners . . . observing and documenting information about the developmental levels of children, participate in meetings to develop Individualized Family Service Plans" (Pickett, 1999, p. 7). Paraprofessionals provide support in many capacities and act as assistants to (a) general and special education teachers, and (b) related servers such as speech language pathology, physical therapy, occupational therapy, adaptive physical education, vision, hearing, and social work. They also work in non-instructional areas, which are transportation, school office, lunchroom, and playground. According to the U.S. Department of Education, paraprofessionals can provide instructional support but they should not be providing instruction or introducing new skills, concepts, or academic content to students. In whatever capacity, they always need to work under the direction of a teacher (DOE, 2001; IDEA, 2004). Although their roles are varied, paraprofessionals can do their responsibilities more efficiently with

adequate training and further education and as a result, perhaps be involved with fewer problems. Individuals who work only in food services, cafeteria or playground supervision, personal care services, non-instructional computer assistance, and similar positions are "not considered paraprofessionals" under U.S Department of Education (NCLB, 2001, Part A (p. 1)).

Statement of the Problem

Paraprofessionals have often had to learn on the job and their pre-training may be minimal. While they may have access to the teacher and his or her knowledge, the teacher is also heavily involved in instruction, paperwork, and team collaboration. The time set aside for paraprofessional training may be insufficient. Therefore, the paraprofessional needs to assume responsibility for becoming trained and educated, not only because of the NCLB requirement, but also to benefit the student with special needs.

The paraprofessional needs to have a working knowledge of specific roles and required responsibilities. This can be found through college courses and training. In their course offerings, many colleges throughout the United States have included classes geared specifically for the paraprofessional. Educators have written textbooks meant to be used in such courses. Most college textbooks have accompanying test banks and problems sets. A test bank can be used to (a) evaluate the students and to assess their learning, (b) serve to motivate and to help them to study more efficiently, (c) help the instructor to understand how successful his or her teaching efforts are in presenting the material, and (d) reinforce learning by providing students with indicators of the unmastered topics (Davis, 1999). College instructors will be able to use the test bank to evaluate student learning and to monitor effective teaching techniques by reviewing

scores from administered tests. Students can benefit from taking tests because test knowledge can help them understand what they have or have not mastered. Effective tests then are useful educational tools because they provide both instructors and students with information about student learning and comprehension.

Statement of the Purpose

Ashbaker and Morgan (2006) wrote a textbook titled *Paraprofessionals in the Classroom* (2006) for paraprofessional education but did not write a test bank. After extensive research on creating test questions and writing a test, a test bank was written for the textbook. The researcher wrote a sample test bank to be field tested and then wrote the final test bank. Before the test bank could be published with the textbook, the questions needed to be field tested to determine if they were effective in assessing paraprofessional knowledge. The purpose of this paper is to show the reliability, validity, and fairness results of the field test created for the textbook. The test bank for this text needs to be an effective tool for paraprofessional education. An effective test bank can show both the paraprofessionals and their instructors what they have learned and what they still need to know. The goal was to provide such an instrument.

Research Questions

Test banks need to be evaluated to determine if they are valid, reliable, and fair. In evaluating *Paraprofessionals in the Classroom* (Ashbaker & Morgan, 2006), the following questions were asked:

1. Did the test bank provide a variety of types of questions?
2. Was the range of difficulty levels sufficient so that the instructor would be able to select applicable items to assess the student knowledge?

3. Were the questions analyzed for item bias, item difficulty, and item-objective congruence?

4. Did the questions have content and construct validity?

CHAPTER TWO

LITERATURE REVIEW

Looking at the Past

As court cases involving paraprofessionals have significantly impacted the field of special education, the training and education of paraprofessionals has become increasingly important for providing education for students with special needs. The first section of this review contains data regarding the number of court cases involving paraprofessionals in the United States between 1994 and 1999. Next, in looking for solutions, what the experts consider as criteria for a valid, reliable, and fair test bank is shown. By writing quality textbooks and corresponding valid, reliable, and fair test banks for college classrooms, the paraprofessional who chooses to become educated with this method will help achieve the goals of NCLB. He or she becomes a highly qualified paraprofessional who can serve students with special needs in an effective way.

Problems with paraprofessionals have arisen because of inadequate training or lack of education. The misunderstanding or neglect of training in legal issues has resulted in law cases. For the benefit of the students with disabilities, NCLB currently requires only Title I paraprofessionals to become more skilled through training or education. By using the knowledge gained from training and education, the paraprofessional can be aware of and responsible for what can and cannot be done with students. This may be a partial solution to the legal issues caused because of lack of understanding roles and responsibilities, which result in court involvement. Table 1 is a review of research done by Katsiyannis, Hodge, and Lanford (2000) showing the number of law cases involving paraprofessionals.

Table 1

Paraeducator-Related Legal Activity by Type and Year, as Reported in IDELR (1994-1999)

Year	Hearings	OCR	OSEP	Court
1999	0	0	0	0
1998	5	3	0	0
1997	7	0	0	4
1996	3	3	0	3
1995	2	5	0	3
1994	5	2	1	1
Total	22	13	1	11

Note. IDELR=Individuals with Disabilities Education Law Reporter; OCR=Office of Civil Rights; OSEP=Office of Special Education Programs.

A dramatic example of the issues resulting from possible improper training was the Armijo case, which resulted in the death of a student. The case involved a *danger creation* situation, which occurs when defendants recklessly create the danger for a student due to neglect. In *Philadelfio C. Armijo-deceased minor; Juanita D. Chaves & Atancio Armijo plantiffs v. Wagon Mound Public Schools* (1998), the paraprofessional had been told by the student that he was going to commit suicide. She did not share this information. This may have been a case of poor judgment, inappropriate confidentiality, or insufficient training. Having received proper training, she would have known that she should have notified school personnel of the student's state of mind. This could have prevented the tragedy.

Looking for Solutions

To prevent such incidents such as *Philadelfio C. Armijo-deceased minor; Juanita D. Chaves & Atancio Armijo plantiffs v. Wagon Mound Public Schools* (1998) from happening, paraprofessionals need more training and education. NCLB requires paraprofessionals to have an associate's degree, pass a stringent test, or show a portfolio of applicable trainings. In order to meet the need for qualified paraprofessionals, many colleges have included classes for paraprofessionals in their curriculum. Because these requirements are relatively recent, this field is growing. Textbooks written expressly for paraprofessional education are coming on the market. One was reviewed for this study, *Paraprofessionals in the Classroom* (Ashbaker & Morgan, 2006). This book explores the roles and responsibilities of paraprofessionals and gives scenarios where the reader can apply chapter information to a situation in which an appropriate interaction is needed between paraprofessionals, professionals, and students. This textbook did not have an accompanying test bank so one needed to be developed and tested. The following section discusses how to create questions and how to test for validity, reliability, and fairness. After extensive research, a test bank using the specific criteria to make a quality test bank will be created.

Creating a Test

The test creator needs to consider issues such as types and levels of questions and how to make them valid, reliable, and fair. He or she needs to write pertinent questions for a particular audience. When the exam is created, it must be checked to see if it can do what it is designed to do. If not, it needs to be rewritten. A summary of Berk's (1984) suggested steps follows. These are needed to create a valid, reliable, and fair test:

1. Create the questions,
2. Select the paraprofessionals,
3. Set up a study to evaluate validity, reliability, and fairness,
4. Field test the material,
5. Review and analyze the data, and
6. Rewrite the material to create the test.

In order to achieve an accurate assessment of students' knowledge, one creates the questions with these concepts in mind.

Create the Questions

A test question should be "important enough to ask and clear enough to answer" (Wormeli, 2006, p. 74). A variety of questions should be used because this gives a better picture of students' mastery. Some students will be able to answer different types of questions better than others will. The quality of a test is based upon the quality of the individual questions; therefore, each question itself needs to be valid, reliable, and fair. Scores obtained from poor tests will have little significance.

There are several different types of questions and each type has a different purpose and level ranging from cognitive knowledge to evaluative level. The instructor needs to choose the type of question based on the desired outcome (Berk, 1984). The intent of the questions may be informative or summative. The instructor may be seeking information regarding the pre- or post-test knowledge level of the students. Table 2, adapted from Berk has different types of questions and has a list of advantages and disadvantages of each.

Table 2

Summary of Different Types of Questions Used for a Survey

Type	Advantages	Disadvantages
Multiple choice	Measures all abilities Wide samples of content Analyzed for effectiveness Efficient scoring	Tendency to measure recall Stems and choices need to be logically and grammatically correlated. Cognitive level
Matching	Easy to write Quick to answer Tests large sample Efficient scoring	Tendency to measure recall All answers written in selections Cognitive level
Fill-in-the-blank	Easy to write Easy to score Efficient scoring	Tests only specific sentences Teacher may use idea instead of text Cognitive level
True-false	Tests large sample of content Efficient scoring	More guessing Difficult to write Cognitive level

Table 2 (continued).

Type	Advantages	Disadvantages
Short answer	More content coverage Less guessing Easy to create	Limited range of abilities Difficult to score Comprehension level
Essay	Easy to create Eliminates guesswork	Limits content Difficult to score because they require judgment from the educator who may have difficulty scoring fairly. Tests higher order thinking such as application, analysis, synthesis, and evaluation

Multiple choice questions. Further analysis of question types reveals more information. For example, special consideration needs to be used when writing multiple choice questions (Berk, 1984). McTighe and Ferrara (1998) have listed nine excellent ways to write multiple choice questions. The last item is included from Survey System (2003) because it points out a concept that may not usually be realized by test creators but are well known to test perceptive students. These 10 concepts are summarized as follows:

1. One question tests only one item,
2. Test understanding, not recognition of terms,
3. Make answer options homogeneous in nature,
4. Make answer options inferior to the real answer,
5. Make the correct answer the same length or shorter than the other options,
6. Make the answer choices grammatically match the question stem,
7. Make the choices logically compatible with the stem,
8. Make the choices plausible,
9. Avoid using *never, always, all,* and
10. Vary the location of the correct answer to avoid one or two consistent letter choices.

Berk (1984) suggests one should first study the structure and content of the sentence and break it into facets. Then, each sentence is mapped so the question stem is obvious. From there, create correct answers and viable wrong answers.

Matching questions. Walker (1998) stated that matching questions test the student's recognition of relationships. This is a good type to use for terms and definitions, cause and effect, parts and units, and problems with solutions. While these provide maximum coverage at a knowledge level for the instructor, they are time-consuming for the student. A student may benefit from guesswork when matching questions and answers, especially if there is an equal number of stems and choices. To write good matching items, one should use 15 items or less and clarify the directions for matching. When using items in the response column more than once, one is able to reduce the effect of guessing. Responses need to be in some type of logical order and should be short. A

matching question usually provides a knowledge level question and a well-written one may give a comprehension level (Bloom, 1956; Walker, 1998).

Fill-in-the-blank questions. Using fill-in-the-blank questions measures recall, which is a knowledge level question, the lowest level on Bloom's taxonomy. A way to make this become comprehension level is to write it as a concept from the chapter instead of having the student recall a sentence (Berk, 1984). This should cause the student to think about what had been learned from the content instead of memorizing sentences.

True-false questions. To create true-false questions, study the construction of the sentence and reverse some part of it so it makes the sentence wrong. One may add "According to the text,…" so the student understands the answer requested is not personal opinion (Davis, 1999). One needs to avoid double negatives and complex sentences and use only one central idea per question. Use exact quantitative words and make more questions false than true since students usually mark more answers true than false (Walker, 1998). True-false questions test knowledge on the lowest level of cognition, which is knowledge (Bloom, 1956).

Short answer questions. By analyzing the sentence, one can create short answer and essay questions. Decide what will help the student achieve a better understanding. Examples of this type are case studies. These can be open-ended questions or visual representations such as charts and graphs (Anderson, 1972). These questions are good for the higher levels of thinking such as application, analysis, evaluation, and synthesis. Writing the question and not the definition tests students' knowledge more deeply. One should also phrase the question so that there is only one answer possible (Walker, 1998).

A summary of Finn's (1978) Anderson's (1972) question writing suggests the following:

1. Replace the anaphoric phrases with actual nouns,
2. Drop extra clauses, and
3. Put a "wh" (what, when, where, why) or how word at the beginning of the sentence stem.

Essay questions. Essay questions come in two types (a) extended response that gives freedom in the answer, and (b) restricted response that gives parameter outlines. When providing students with essay questions, one needs to provide a limit to the amount of time for providing the answer. Use a definitive task word such as compare, analyze, or evaluate to help the student understand the question. Essay questions show the instructor how the student arrived at the answer. Application, synthesis, and evaluation level questions can be asked using essay questions (Bloom, 1956; Walker, 1998).

Levels of Questions

As well as the types of questions, one must consider the levels of questions. As discussed in the preceding section, one can use several different levels, depending upon what one is seeking. Bloom (1956) created the well-known taxonomy of levels of knowledge. In addition, Furhman & Grasha (1983) have listed specific words to use in writing the various levels of questions. They are (a) cognitive level uses words such as define, describe, identify, label, list, match, name, outline, reproduce, select, and state; (b) comprehension level which uses convert, defend, distinguish, estimate, extend, generalize, give examples, infer, predict, and summarize; (c) application level has demonstrate, modify, operate, prepare, produce, relate, show, solve, and use; (d) analysis level uses diagram, differentiate, distinguish, illustrate, infer, point out, relate, select,

separate, and subdivide; (e) synthesis levels utilize categorize, combine, compile, devise, explain, generate, organize, plan, rearrange, reconstruct, revise, and tell; (f) and evaluate questions can contain appraise, compare, conclude, contrast, criticize, describe, discriminate, explain, justify, interpret, and support. By following these guidelines, the test creator can not only provide a variety of questions but also use the results to establish the learning level of the students.

Placement of questions and choices is important. A summary of Wormeli"s (2006) suggestions are as follows:

1. Keep matching items on the same page,
2. Provide T and F for students to circle for true-false questions so one does not need to try to interpret which letter the student may be trying to write,
3. Keep fill-in-the-blank spots close to the end or stem of the question so reading comprehension does not become an issue,
4. Highlight key words, and
5. Clarify for the students what is expected.

Wormeli (2006) also suggests that several short exams are better than one long one because they will give better snapshots of the students' progress.

Selection of Paraprofessionals

The selection of the paraprofessionals depends upon several criteria (a) the availability of the paraprofessionals, (b) the time they require to assess the test questions, and (c) their interest and abilities. One can choose veterans or novices to show how difficult the test could be. It is assumed that it would be easier for an individual who is considered a veteran in that field to understand the test than one who is not (Berk, 1986).

Validity, Reliability, and Fairness

In addition to the above criteria, one needs to consider validity, reliability, and fairness of the test. Validity is the measurement of how well an assessment actually measures what it is supposed to measure. Reliability shows that the scores would be consistent over time or across evaluators. Fairness promotes the assessment of the test being equally fair regardless of gender or background.

The purpose of validity is to prevent unforeseeable and negative results. These consequences might include item-objective congruence, item bias, and item difficulty (Berk, 1984). Content and construct validity are also used to measure effective test questions (Messick, 1995). Berk states that item-objective congruence is checking to see if the question comes from the actual written content that the student has studied. It would not be appropriate to ask a question if the answer was not in the material. Congruence also is reflected in the verb and construction of the question. If the question is posed as singular and only one of the choices is singular, the answer is too obvious. Item bias is created when one does not consider (a) offensive gender and cultural or racial stereotyping, (b) offensive language, (c) activities and words not familiar to the students and (d) items not normed for the particular population taking the test (Hambleton, 1980). Researchers from Learning Point Associates at North Central Regional Education Laboratory (2005) state that item bias generally refers to a group of individuals rather than to a single individual. Item difficulty is overcome when the test starts out with the easier questions at the beginning and the more difficult ones at the end. The purpose is to let all students have some success yet present a challenge for the more advanced students as they finish the test. Berk suggests this format because it evaluates all levels of readers.

Content validity is the extent to which the content of the test represents a sufficient sampling of the knowledge and skills taught to the class. Davis (1999) states that the test questions should cover the concepts in proportion to the emphasis they received in the class. Since the instructor cannot assess the students on every concept presented in the classroom, he or she must select questions that represent the main ideas that have been presented. Therefore, the questions need to reflect not only the proportion of the emphasis but also cover the variety of topics covered (Cronbach, 1971).

Messick (1995) states that construct validity are the extent to which a test measures what it is supposed to measure. Construct validity can be done by using the contrasting groups approach. For example, comparing a single idea by using individuals from different domains would be assessing construct validity. An expert in the field usually measures this. Another way to improve construct validity is to select paraprofessionals who have no relationship with the test creator so that they do not try to answer how they believe the creator would like them to do. This is called the Hawthorne effect. In addition, when an examiner is free to not complete a test because a grade is not required, construct validity is not affected by evaluator apprehension (Messick). Concepts used to analyze construct validity are (a) clear test questions, (b) understandable instructions on the answer sheet, (c) similar test conditions for all paraprofessionals, (d) motivated paraprofessionals who perform to their best ability, (e) interesting content, (f) suitable vocabulary, (g) examiner ability, and (g) suitable time of day (Berk, 1984).

Trochim (2005) teaches that reliability is shown when the scores are consistent across evaluators or over time. The results should not be affected by when the assessment occurs or who scores the test. Reliability shows that the measure used would give the

same result if it were used repeatedly. Reliability can be affected by the testing threat of a test-retest style assessment. For example, if one needed to pre-assess the knowledge level of a group of students by giving them a pretest, he or she could not use the same test to assess the pre-knowledge level again because the students would have learned from the first assessment. The parallel forms method is used to rate reliability. This is creating a large group of questions that address the same construct (Trochim).

Fairness has to do with equality. According to the NWREL (2001), the assessment scores should not be influenced by gender or background of paraprofessionals. All students need to have the same opportunity to take the exam with the same accommodations. In order for this to occur, one could have the exam take place at the same time in the same place for all students. This is a common practice in many classrooms. Bias is different from fairness because the former may refer to a group of individuals and the latter may refer to a single individual. The content of the test needs to reflect commonalities among students, not differences. For example, if an exam were considered fair, an instructor could ask the students to put a number on the exam instead of names. Then, as they were being graded, the instructor would not be influenced by subconscious or conscious attitudes toward any one particular individual.

Field Test the Material

In preparing a test bank, the material needs to be field tested before it is presented to the public. This is to find errors that have been created or missed. The test can first be done with a pre-survey that will help to find problems with such items as instruction clarity, length of test, question and answer congruence, bias, and other pitfalls of test creation (Berk, 1986).

Reviewing and Analyzing the Data

The purpose of analyzing the data is to find the response patterns, which will help to reveal question errors. When the pre-survey is analyzed and the data used to write appropriate questions, the material needs to be field tested in order to cover the issues referred to in previous sections (Berk, 1986). Using the data collected from the field test will help to create a test bank or exam that will have a good reputation and can be used at the appropriate times and with the correct populations. Each piece of data is an observation which, when collected, will create classes of information. The frequency of the classes can show the students' understanding of the questions asked. The instructor can create a histogram to show the data in a graph form. He or she can also discover the mean, median, and mode of the answers and use this to correct his or her teaching style (Neely, 2004). If the instructor finds that the mode of the scores is far below where is expected, a change of teaching style may need to occur.

Rewrite the Material to Create the Test

The test creator needs to decide what type of test is needed and select applicable questions. The most common choices are the proficiency test and the achievement test which can be used as a pre-test and the achievement test as a posttest, respectively (Frost, 2002). After data results, the test creator revisits the questions and makes changes needed to make the test more valid. Test questions more understandable for the student and the results more useful for the instructor. A summary of Frost's suggestions follows:

1. Decide what the purpose of the test is,
2. Make a list of what the test needs to cover,
3. Consider the length and layout,

4. Weight the questions according to their importance,
5. Write the questions,
6. Write the instructions,
7. Decide on the grades, and
8. Create a grading scale for essay and short answer questions.

By following these guidelines, a test can be created which will not only assist the instructor in assessing the students' knowledge level but can also help in showing the strengths and areas in which the teacher needs to improve in teaching or creating tests.

CHAPTER THREE

METHOD

This chapter explains the various aspects of the method used to create the field test. First, is the discussion the participants, the criteria used to select them, and their involvement. Next, the materials used in the field test are explained. As important as these two criteria is the setting in which the testing took place, regarding day, time of day, and place. Then, the research design, which is a test-retest format, is reviewed. Finally, the procedure used to accomplish the examination was discussed.

Participants

The selection of the participants was carefully considered so that the survey would be as valid as possible. Two individuals who were asked to mentor the sessions randomly chose them. Their only criteria required was that they needed to be involved in special education and live in the county in which the research was being done. A wide contrast of paraprofessionals was used.

Validity

To create construct validity, one uses the contrasting groups approach, choosing individuals who do not have the same background. The paraprofessionals who participated were familiar with their own area of special education related services but each area of expertise was different. They were from. Some could have been considered veterans because of their many years of experience. The novices had worked as paraprofessionals for only a short period. The age range was 43 years.

Construct validity was accomplished by the fact that the paraprofessionals had no relationship to the researcher. This avoided the Hawthorne effect when individuals

unknowingly try to provide the answer they believe the test provider wants to have. The researcher selected mentors. They were to choose paraprofessionals to complete the surveys from speech and language, occupational therapy, Title I, early intervention, special education preschool, vision, elementary school, and middle school. After the criteria that were needed for the field test were explained, the mentors selected the participants. Other criteria of construct validity were supported because the paraprofessionals were not required to finish the field test and did not experience evaluator apprehension, an issue stated in the literature review.

Other aspects of construct validity were taken into consideration when (a) the mentors gave clear understandable instructions, (b) the test conditions were similar for all participants, (c) monetary reimbursement was provided to motivate the paraprofessionals, (d) the content should have been interesting to them because it was written for paraprofessionals, and (e) the time of day was suitable because it was not during working hours, thus causing distractions.

Reliability

Reliability, which should occur when assessing, was done with a group of paraprofessionals, not just one individual. When the scores are consistent across evaluators, they are more reliable. The questions which were not consistently answered by the paraprofessionals were changed or deleted. Eight paraprofessionals were invited to do the field testing. The data were collected from surveys and each answer and rating was tabulated to find patterns of poor question writing shown by consistently wrong answers and levels of difficulty, which were marked for each question.

Fairness

Fairness was accomplished because the researcher did not know the gender and the background of the paraprofessionals. However, because of sites chosen, their county of employment and general background were known.

Paraprofessional Demographics

The paraprofessionals gave information regarding ages, years of service, gender, ethnicities, and educational levels. Eight paraprofessionals were chosen from the following areas: speech and language, occupational therapy, Title I, early intervention, special education preschool, vision, elementary school, and middle school. Table 3 is a summary of other demographics.

Materials

The textbook used was Paraprofessionals in the Classroom (Ashbaker and Morgan, 2006) which was written for paraprofessionals who take college courses. It contains 5 sections made of 13 chapters. These sections include (a) background and context, (b) learning environment, (c) instruction, (d) professionalism, and (e) appendix. Some of the topics discussed are definitions, the roles and responsibilities of a paraprofessional, and the United States educational system. In addition, the relationships between paraprofessionals and other adults, how a student learns, how to use effective instruction, how to manage time, how to manage student behavior and classroom and the meaning of ethics and the Council for Exceptional Children (CEC) standards.

The chapter selections were divided so that each examiner received three different chapters and no paraprofessional had the same three chapters as any other. This assisted with creating fair questions. These assignments are shown in Table 4.

Table 3

Paraprofessional Demographics

Respondent	Age	Gender	Years of experience	Ethnicity	Date of survey	Time used for three chapters
1	60	F	8	White	4-18-05	1 hr 32 min
2	61	F	15	White	4-18-05	1 hr 12 min
3	18	F	1	White	4-18-05	1 hr
4	52	F	3 mo	White	4-18-05	1 hr 15 min
5	47	F	10	White	4-18-05	45 min
6	23	F	1.5	White	4-18-05	1 hr 10 min
8	24	M	5	White	4-25-05	55 min
Mean	40.75	85% F	5 yr 11 m			67 min
Median	47		5 yr			70 min

Setting

Six of the paraprofessionals were asked to come to the school district room on the same day at the same time to take the tests. The seventh examiner was asked to take the survey in the classroom where she worked under the supervision of the mentor. The eighth examiner was held in abeyance in case one of the seven was unable to complete the assignment or a survey was done incorrectly. As the later actually did happen, the eighth paraprofessional did that particular set of chapters one week after the original date on the same day and at the same time. The seventh paraprofessional was given the surveys and took them home. This paraprofessional was not monitored during the field test and the test and the data were invalid and therefore not used. The eighth

Table 4

Chapter Assignments

Paraprofessional	Chapter number to be surveyed
A	1, 3,4
B	3, 4, 5
C	4, 5, 7
D	5, 7, 8
E	7, 8, 10
F	8, 10, 1
G	10, 1, 3

paraprofessional was monitored by the mentor from the first school district but the survey was done at the paraprofessional's place of work.

The time chosen was 4:00 p.m. when all the paraprofessionals were finished with their school jobs. The day of the week chosen was a Monday. The rooms for the survey were at the district offices and early intervention site. There was no limitation set on time to complete the questions. This was a way to decrease examiner apprehension, which increases construct validity. The length of time to read each set of sections and answer the questions depended upon the reading ability of the individual person but was estimated from the pre-survey to be 30 to 60 minutes for all three chapters. The actual length of time was between 45 minutes and 1 hour and 45 minutes, which would be between 15 minutes and 35 minutes for each section. This was within the expected time.

Research Design and Procedure

Research was done on various research designs and which one would best fit the survey. The paraprofessionals would not be reviewing their answers or learning from their mistakes. However, this was not the focus, which was on the evaluation of the questions so the design chosen worked well for this situation.

Design

The research design used was a test-retest format. This was an appropriate choice of survey type because there was no opportunity for the examiner to learn from the test or textbook beforehand so the answers were genuine. The first step for the researcher was to write the questions to be field tested from the chapters selected from the textbook. Seven chapters were read specifically for key words and new information for a paraprofessional. These concepts were used to write the questions. Then, the researcher analyzed each question by checking it against criteria from the 2001 Publication Manual of the American Psychological Association (APA). The following items were considered (a) continuity in words, (b) correct punctuation, (c) verb tense agreement (d) redundancy, (e) hard to understand technical terms, (f) wordiness, (g) colloquial expressions, (h) correct grammar, and (i) bias of gender, ethnicity, disability, or age. The test questions were also written so that there would be variety of knowledge levels such as knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956).

Clearance from the Instructional Review Board for Human Subjects (IRB) was received. The IRB proposal is Appendix. Next, the special education directors of two school districts and one early intervention program were contacted in order to obtain permission to have paraprofessionals do the field testing. After they gave permission, two

employees of the school districts were asked to serve as mentors. One of these mentors also worked in the early intervention program. They were asked to find paraprofessionals from eight different areas of special education, set up testing times and places, and mentor them during the test. The mentors contacted the principals of the schools and requested permission to have the paraprofessionals do the surveys. Then, a meeting time and place was set up. Each paraprofessional received a letter explaining the project. It is Appendix B. By having paraprofessionals meet together under the supervision of the mentors, they had no interruptions, no phones calls, and were able to finish the assignments before they left. They read the sections taken from the textbook, turned the pages, read each test question, marked or wrote an answer, and then rated the questions as easy, average, or difficult. The questions had two parts. The first part required an answer from the paraprofessional, which was to be taken from section content. The second part was a rating scale of the question itself. There were three choices: easy, average, or difficult. Three separate individuals evaluated each question three times. Consequently, evaluating all 411 questions gave 1,233 ratings.

The purpose of having the paraprofessionals answer the question itself was not to test their ability, but to encourage them to read the questions carefully before marking the difficulty level. Therefore, the data requested were the opinion of the paraprofessional regarding the difficulty level of each question. There were a varied number of sections so amount of questions for each chapter depended on its length.

Revisions, which were necessary for the rewriting the test banks were made, again, using the criterion researched from the APA manual and reviewing the input from

the paraprofessionals' survey data. The test questions were revised and then used as models for writing the rest of the test bank, which covered five more chapters.

Peer Review and Pre-survey

To search for items such as unfamiliar words, item bias, item objective congruence, content validity, construct validity, clarity of instruction, all of which are requisites stated by the experts in the literature review, a peer read the sections and questions (Berk, 1984; Davis, 1999; Hambleton, 1980; Messick, 1975). As a pre-survey, one paraprofessional assigned by a mentor was asked to review one chapter and its questions for clarity of instruction and use of unfamiliar vocabulary. This is a concept of construct validity.

Procedure

Permission was requested for paraprofessional participation from the directors of two school districts and the director of an early intervention program. It was given over the phone. The supervisors and principals of the paraprofessionals were contacted and gave verbal permission for their paraprofessionals to participate in a survey. Two mentors were chosen. They, in turn, selected eight paraprofessionals to participate based on the criteria of availability, dependability, and domain. These individuals' identities were unknown to achieve exempt status from the IRB.

The mentors were given eight letters to be sent to the paraprofessionals. Each participant was contacted by a cover letter requesting participation in this study. The letter contained the title of the survey, a short introduction telling about (a) the researcher, (b) the reason for the study, (c) the notification of confidentiality, (d) an invitation to take the survey, (e) an explanation of the incentive, (f) how to complete the survey, (g) how to

submit the survey, and (h) how to submit questions or comments. The chapters were randomly selected to cover various topics from the textbook. Because some paraprofessionals are currently employed and others are only in training programs, chapters from the beginning, middle, and end were chosen. Information from the beginning of the text may have been familiar to those paraprofessionals who were already employed. This would make the text fair. Seven of the 13 chapters were selected for assessment. Another criterion was to look for new terminology and key ideas. These could be focal points for an exam. A three point rating scale was provided to have the paraprofessionals mark the difficulty of the questions. They were marked easy, average, or difficult. The survey was explained to the mentors. A sample question was provided with an example of how to mark the answer. Mentors were to give oral instructions and write them on a whiteboard. The paraprofessionals were to read the sections silently and then answer each set of dual questions.

The paraprofessionals' job was to read and to evaluate the questions from the written copies provided by marking the answers, rating the questions, and turning the tests into the monitor. The mentor's job was to pass out the tests, receive them when finished, and pay the participants for the number of tests accomplished. The participants were paid $2.00 for the first chapter completed, $3.00 for the second chapter, and $5.00 for the third chapter. Since there was only a minimum of eight individuals needed and each was paid a maximum of $10.00, the costs were covered by the researcher. A participant who chose not to finish all three chapters was paid for work done. Another paraprofessional would have been asked to participate so that each chapter has three reviews. All eight participants chose to finish all three chapters given to them.

The researcher evaluated the criteria for the question difficulty as shown in Table 5. The ratings were done by the first, second, and third paraprofessional who did the surveys. The ratings were tallied for each question. The questions were read again for criteria from the APA manual and the final test bank was made. The results are what the researcher did with the questions.

Table 5

Question Rating Criteria

First rating	Second rating	Third rating	Results
Difficult	Difficult	Difficult	Deleted
Easy	Average	Difficult	Average
Easy	Easy	Average	Easy
Easy	Average	Average	Average
Easy	Difficult	Difficult	Difficult
Easy	Easy	Easy	Deleted
Average	Average	Difficult	Average
Easy	Easy	Difficult	Easy
Average	Difficult	Difficult	Difficult
Average	Average	Average	Average

CHAPTER FOUR

RESULTS

After creating a test bank for the paraprofessional textbook, *Paraprofessionals in the Classroom* (Ashbaker & Morgan, 2006), it was field tested by eight paraprofessionals from eight different special education related service areas. The paraprofessionals read a selection, answered the questions, and rated the questions for difficulty. The data were analyzed. This chapter provides the results of this analysis.

Reviewing the Research Questions

Test surveys were evaluated to determine if they were valid, reliable, and fair. In evaluating *Paraprofessionals in the Classroom* (2006), the following questions were asked and the results to each question are provided below:

1. Did the text bank provide a variety of types of questions?
2. Was the range of difficulty levels sufficient so that the instructor would be able to select applicable items to assess the student knowledge?
3. Were the questions analyzed for item bias, item difficulty, and item objective congruence?
4. Did the questions have content and construct validity?

Variety of Questions and Levels of Difficulty

The test bank contained a variety of questions. These were multiple-choice, true-false, fill-in-the blank, short answer, essay, and matching. The variety provided information regarding the different levels of cognitive and higher level thinking. This gives the instructor an opportunity to find individual data for each student. Table 6 is a comparison of the difficulty of questions marked according to chapter topic.

Table 6

Survey Results

Chapter Number	Chapter Content	Percentage Marked		
		Easy	Average	Difficult
1	What is a paraprofessional?	39	61	0
3	Paraprofessional duties	35	58	7
4	Paraprofessional standards	33	62	5
5	Organization and management of the learner	40	33	27
7	Instructional supervision	32	68	0
8	Effective instruction	30	60	10
10	Time management	47	53	0
Mean		36.5	56.45	7

Note. Chapter 8 is significantly longer than the other chapters.

After tallying the data, questions were revised, kept, or deleted from each chapter. This was done by following the guidelines in the APA manual. Table 7 is a display of revisions per chapter. Based on the types of questions, Table 8 follows and is the comparison of the levels of knowledge using easy, average, and difficult.

Validity

Item-objective congruence. After reviewing the data, all questions for item-objective congruence were read to see if the question actually came from the text selections. It would not have been appropriate to ask a question if the answer was not in the material. Those reviewed were marked as difficult to see if the question actually was in the preceding content. Congruence also was reflected in the verb and construction of the selections of answers. If one was singular, they both needed to be.

Table 7

Question Revisions in Percentages per Chapter

Chapter	Rewritten	Same	Deleted
1	57	38	5
3	39	26	35
4	15	56	29
5	6	94	0
7	19	56	25
8	45	55	0
10	27	70	3

Item bias. Two questions had to be rewritten for item bias. In both of these, the pronoun “he” had to be changed to “he and she”. Item bias can become an issue when one fails to consider (a) offensive gender and cultural or racial stereotyping, (b) offensive language, (c) activities and words not familiar to the students and (d) items not normed for the particular population taking the test. Because the textbook was written for

Table 8

Question Types Compared to Knowledge Levels

Question Type	Knowledge Level	Difficulty Rating in Percentages		
		Easy	Average	Difficult
Multiple Choice	Cognitive	74	21	5
Matching	Cognitive	36	42	22
Fill-in-the blank	Cognitive	8	14	78
True-false	Cognitive	39	42	19
Short answer	Comprehension	45	32	23
Essay	Application, analysis, synthesis, evaluation	11	29	60

paraprofessionals, the activities and the words may have been familiar to them.

Item difficulty. One may have thought that the paraprofessionals would mark the cognitive level questions as easy and the higher level thinking ones as difficult, but this was not the case. They did have a tendency to mark the short answer and essay questions as easy, possibly because these required an opinion rather than a fact. However, the answers were not always correct.

Content validity. The questions were reviewed for content validity by rereading the sections and checking to see if they written were from the main points or key words.

All questions met this requirement. The other aspect of content validity was to have the questions match the emphasis used in the classroom. For example, a topic not discussed in the classroom should not have a question written about it in a test. Since the sections only were read once and not taught in any other manner, this criterion using the data from classroom emphasis was not available for use.

Construct validity. Most of construct validity has to do with paraprofessional performance. One way to create construct validity focused on using contrasting paraprofessionals from the following areas (a) different special education fields, (b) veterans and novices, and (c) individuals of different years of experience and ages. Another construct validity criterion was accomplished because the paraprofessionals had no relationship to the researcher. This avoided the Hawthorne effect when paraprofessionals unknowingly try to provide the answer they believe the test creator wants to have. The mentors selected the participants after an explanation of the criteria needed for the field test. Another criterion of construct validity was supported because the paraprofessionals were not required to finish the field test and did not experience evaluator apprehension, an issue stated in the literature review. After reviewing the information provided on the level of difficulty, the variety of question types, the validity, reliability, and fairness of the questions, the test questions were revised, deleted, or kept. Using these as models, questions were written for the remaining chapters, maintaining a percentage of 50% cognitive level questions and 50% short answer and essay question for each chapter. The chapters started with easy questions and ended with more difficult questions, as suggested by the experts in the literature review.

CHAPTER FIVE

DISCUSSION

The first research question was asked to see if the test bank provided a variety of types of questions. Some of each of the different question types in the literature review are included. They are multiple-choice, matching, fill-in-the-blank, true-false, short answer, and essay. An example of each of these types of questions is in Appendix C. The questions follow both the guidelines from Bloom (1956) and Furhman and Graha (1986) who use specific words to help the student understand exactly what is expected as an answer.

The second question dealt with the range of difficulty levels. These were analyzed to see if the instructor would be able to select a range of items to assess the student knowledge. After reviewing the data from the examiner ratings, the survey questions were rewritten to have 50% of the questions on the cognitive knowledge level, using multiple-choice, matching, Fill-in-the-blank, and true-false questions. These are the types of questions most often seen in pretests because they are the easiest to write, according to Berk (1984). Questions better suited for a posttest are short answer and essay which cover comprehension, application, analysis, synthesis, and evaluation (Berk). Multiple-choice, matching, fill-in-the-blank, and true-false questions are easier to correct, short answer and essay questions provide the instructor with more knowledge about the students' progress because they require the higher level of thinking.

The purpose of research question three is to search for item bias, item difficulty, and item-objective congruence. The purpose of the pre-survey done by the first paraprofessional was a search for item bias and all questions were analyzed for item bias

after the final data had been collected. Item difficulty was avoided when the cognitive knowledge level questions, multiple-choice, matching, true-false, and fill-in-the-blank, were put at the beginning of each chapter test. This is to give each student some change of success. The more difficult questions, short answer and essay, are put at the end of the chapter test. This is to challenge the more advanced student. Item-objective congruence was reviewed both before and after the field testing. All questions marked three times as difficult or three times as easy were deleted. The question may have been stated awkwardly. There were no item-congruence issues with multiple choice or matching questions. For example, if the subject was singular then all the choices needed to be singular.

The purpose of the final question was to seek for content and construct validity. The peer from a university class who did the first review analyzed the sections and questions for content validity. The sections and questions were reviewed again after the field testing by reading the topic sentences of the paragraphs, checking for key words, and new vocabulary. The questions were created from these. In the literature review, Messick (1995) wrote that construct validity is included by the following (a) contrasting groups, (b) no relationship between paraprofessionals and researcher, (c) no evaluation apprehension, (d) interesting content, and (e) suitable vocabulary. These conditions were all present during the survey sessions.

By analyzing the data through the peer survey, the pre-survey, the field testing and the pre- and post-review, the work was assessed for all criteria listed (a) different types and levels of question difficulty, (b) validity, (c) reliability, and (d) fairness. Following the recommended procedures from research done on testing which was

included in the literature review, the questions were written, edited, and deleted. By doing this, the test bank was ready for use by instructors in the field of paraprofessional education.

Benefits and Limitations

The lack of paraprofessional education and training is a significant issue for the educational field because it may affect the quality of services delivered to students with whom the paraprofessionals work. It is a purpose of NCLB to have this training and education focused on students with special needs. It is also a great benefit to have paraprofessionals well trained, well educated, and justly confidence in laws, rules, and regulations. It is important to have a paraprofessional who does not make mistakes.

Benefits. One of the benefits of taking a college course while working as a paraprofessional is that one can apply the knowledge learned in the class on the job. Under the direction of an educator, a well-trained paraprofessional can assist in making the student's education appropriate and specific to individual needs. Although good experience can be a great teacher, formal training and education create a more highly qualified paraprofessional. By using textbooks written specifically for paraprofessional education and corresponding valid test banks, instructors are able to educate paraprofessionals so they not only know what they are supposed to do but how to do it, also. When paraprofessionals take an exam and answer questions inaccurately, they need to review the exam and find the correct information. Another benefit of this study was the variety of the areas of special education in which paraprofessionals worked. This helped to make the study broader because they had different views of what was important to learn.

Limitations. A limitation of this study was the small number of paraprofessionals who participated in the study. However, the focus was on the difficulty of questions, not on the knowledge of the paraprofessionals. It was also a limitation of the study that all respondents are White and 85% are women. However, according to Pickett (1999), 95% of the paraprofessional force is women so this limitation was actually more representative of reality than if the genders had been equally represented. Another limitation was that the area of special education represented by the paraprofessional who was given inaccurate instructions was not represented.

Implications for the Future

As seen by the death of a young student in *Philadelfio C. Armijo-deceased minor; Juanita D. Chaves & Atancio Armijo plantiffs v. Wagon Mound Public Schools* (1998), the lack of training can make a devastating difference. The education and training of the paraprofessional is not to be taken lightly or brushed off as unnecessary. The knowledge needed to become a highly qualified paraprofessional as required by NCLB can be obtained in the college classroom. Using textbooks and test banks as a method to teach paraprofessionals is an excellent way to educate them. Valid, reliable, fair test banks show the level of the paraprofessional's current knowledge and deficits. Comprehensive, understandable textbooks will guide the instructor in addressing knowledge deficits. The extreme variety of skills necessary requires a comprehensive knowledge based on the same concepts required for educators. Paraprofessionals training needs and educational deficits need to be addressed in a manner that will facilitate an appropriate education for students with special needs. This can be one way no child will be left behind.

Future research could be conducted on the effect of paraprofessional training education on students with special needs before and after the implementation of NCLB. Some aspects to consider are the increase or lack of progress made by students. Once the paraprofessionals have become highly qualified, research to see if students of those paraprofessionals who are highly qualified according to NCLB standards progress further than those who are not highly qualified. Other data to track would be to see if paraprofessionals go on to become highly qualified teachers because they have continued their own education.

REFERENCES

American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (5th ed). Washington, DC: Author.

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42,* 145-170.

Armijo By and Through Chavez v. Wagon Mound Public Schools, 159 F.3d 1253 (10th Cir. 1998).

Ashbaker, B. Y. & Morgan, J. (2006). *Paraprofessionals in the Classroom*. Boston: Allyn & Bacon.

Berk, R. A. (1986). *Performance assessment: Methods and applications*. Baltimore, MD: John Hopkins University Press.

Berk, R. A. (1984). *Guide to criterion-referenced test construction.* Baltimore, MD: John Hopkins University Press.

Bloom, B. S. (1956). *Taxonomy of educational objectives. 1: Cognitive domain*. New York: McKay.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). Educational Measurement (2nd ed.) Washington, D.C.: American Council on Education.

Davis, B. G. (1999). *Tools for teaching*. University of California-Berkeley: Jossey-Bass. Retrieved October 24, 2005 from http://honolulu.hawaii.edu/

Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement, 32*, 255-265.

Frost, R. (2002). *Test writing.* Retrieved July 15, 2005 from http://www.teachingenglish.org.uk/about.shtml

Fuhrman, B. S. & Grasha, A. F. (1983). *A practical handbook for college teachers.* Boston: Little, Brown.

Giangreco, M. F., Yuan, S., McKenzie, B., Cameron, P., & Fialka, J. (May/June 2005). Be careful what you wish for . . . Five reasons to be concerned about the assignment of individual paraprofessionals. *Teaching Exceptional Children 37*(5), 28-34.

Hambleton, R. K. (April, 1980) Review methods for criterion-referenced test items. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C.

Katsiyannis, A., Hodge, J., & Lanford, A. (September-October, 2000). Paraeducators: Legal and practice considerations. *Remedial and Special Education (21)*5, 297-304.

Learning Point Associates. *A Tool Kit for Professional Developers: Alternative Assessment.* North Central Regional Educational Laboratory. Retrieved October 13, 2005 from http://www.ncrel.org/sdrs/areas/issues/methods/assment/as5relia.htm

McTighe, J. & Ferrara, S. (1998). *Assessing Learning in the Classroom*. Washington, D. C.: National Education Association of the United States.

Messick S., (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955-66.

Neely, J. (2004). Data Analysis. University of Northern Iowa. Retrieved October 13, 2005 from http://fp.uni.edu

Pickett, A. L., (1999). *Strengthening and supporting teacher/provider-paraeducator teams: Guidelines for paraeducator roles, supervision, and preparation.* City University of New York: National Resource Center for Paraprofessionals in Education and Related Services, Center for Advanced Study in Education Graduate Center.

Survey System. (2003). Creative Research Systems. Retrieved December 27, 2005 from http://www.surveysystem.com/sdesign.htm.

Trochim, W. M. (2005). *The Research Methods Knowledge Base*, (2nd ed.). Retrieved December 29, 2005 from http://trochim.human.cornell.edu/kb/index.htm

U.S. Department of Education. *No Child Left Behind Act* (2001). Title I Paraprofessionals Non-Regulatory Guidance. [Electronic version]. Retrieved February 1, 2005 from http://Ed.gov/policy/elsec/guid/paraguidance.doc

Walker, G. H. (1998). *Designing Test Questions*. University of Tennessee at Chattanooga. Retrieved October 13, 2005 from http://utc.edu/

Wormeli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom.* Portland, ME: Stenhouse Publishers. Retrieved April 2, 2006 from http://www.stenhouse.com/pdfs/

APPENDIXES

Appendix A

Synopsis of Proposal February 28, 2005

1. Specific Aims

   The specific aim of this study is to create a test bank for a textbook written for paraprofessional education.

2. Hypothesis

   Because a sample of the test bank will be field tested on a variety of paraprofessionals, the data collected from the survey will be applicable in assisting in the creation of valid, pertinent, and fair questions.

3. Background and Significance

   As of January 2006, all paraprofessionals working in a Title 1 school are required to pass a stringent state test, have an associate's degree, or present a portfolio in order to continue their employment. In addition, paraprofessionals need training in legal issues, behavior strategies, child characteristics, and teamwork concepts. The test bank is created to align with a textbook written expressly for paraprofessional education in a college classroom.

4. Description of Subjects

   The eight paraprofessionals who will be requested to participate in the study will be from the fields of early childhood special education, early childhood intervention, speech and language, vision, occupational therapy, physical therapy, office staffing, and middle school special education. They are above 18 years old and work for Alpine School District, Nebo School District, or Kids on the Move.

5. Confidentiality

A monitor will be contacted who will stay with the paraprofessionals as they complete the survey. The researcher will not attend this nor will she know who the participants are, other than that they are paraprofessionals and are from a variety of fields. The surveys will be collected from the participants. The surveys will specifically request that no name be written. Information requested will be years of experience, age, gender, area of special education, time started, and time finished. The surveys will be kept in a locked filing cabinet after the data has been analyzed.

6. Method or Procedures

The textbook was read and extensive research on how to write valid test questions was done. Next, sections of the text were selected, the questions written. Then, the special education department supervisors and early childhood intervention director will be contacted for permission to contact paraprofessionals. Then, supervisors of various fields of education will be contacted and requested that they give a letter to one of their paraprofessionals who would be willing to participate in the survey. A district person will be requested to monitor a survey session at a district room where the paraprofessionals can complete the survey at the same time. The monitor will contact the paraprofessionals to remind them of the survey session time. The survey will take place and at the end as each paraprofessional finishes, she or he will turn in the survey to the monitor and receive compensation. The surveys will be analyzed. The data will be used to revise the test questions. The revised test bank will be submitted to Allyn and

Bacon Publishers.

7. Data Analysis

Each chapter will have a varied amount of sections and correlating questions due to the length of the chapters. Each paraprofessional will be given three chapters to read, answer the question, and rate the difficulty of the question. Each chapter will be read three times by three different paraprofessionals.

Paraprofessional A will read chapters 1, 3, 4

Paraprofessional B will read chapters 3, 4, 5

Paraprofessional C will read chapters 4, 5, 7

Paraprofessional D will read chapters 5, 7, 8

Paraprofessional E will read chapters 7, 8, 10

Paraprofessional F will read chapters 8, 10, 1

Paraprofessional G will read chapters, 10, 1, 3

An eighth paraprofessional will attend the session and do the survey if any one of the other paraprofessionals has to leave early or do not show up so that every chapter is read three times.

The questions will be evaluated according to two criteria (a) was the question answered correctly and (b) what was the rating given the question.

8. Risks

Some people experience test anxiety. Because it is the questions that are being evaluated and not the paraprofessional's knowledge level and it is voluntary, there should be little if any test anxiety.

9. Benefits

The benefits will be that the questions will be field tested by the skill level of person who would be actually using the text and questions in a college classroom. A benefit to the paraprofessional is that he or she will have a preview of what will be in the text and a little more knowledge of the roles and responsibilities of the paraprofessional job.

10. Compensation

The paraprofessionals will be paid $2.00 for the first chapter completed, $3.00 for the second chapter, and $5.00 for the third chapter. This is to encourage the completion of all three chapters. The eight paraprofessionals' costs will total $80.00. No funding is sought for this.

Appendix B

Cover Letter to Paraprofessionals

"Informed Consent Statement" for an "*Exempt*" Research Survey

This survey is being conducted by Linda Hansen, a Brigham Young University graduate student, to determine the effectiveness of textbook questions written for an unpublished textbook by Dr. Betty Ashbaker and Dr. Jill Morgan for paraprofessional education. The paraprofessional will read up to three chapters' selected sections, answer the questions, evaluate the questions, and submit the survey. Each chapter completed will generate $2.00 which will be given to the paraprofessional.

Participants will be chosen randomly from the Alpine, Nebo, and Kids On the Move by contacting first the director of special education, then the principal or director to obtain consent for participation.

The survey consists of 4 to 14 sections and will take 15 to 30 minutes, depending upon the chapter chosen and the reading ability of the participant.

There are minimal risks or and/or benefits to your participation in this study. The risk is that the paraprofessional might experience discouragement because the questions may not be understandable. The benefit is that the paraprofessional is contributing to the validity of the text questions for a textbook that will be used in the future to educate paraprofessionals in their educational classes.

Involvement in this research project is voluntary. You may withdraw at any time without penalty or refuse to participate entirely. There will be no reference to your identification at any point in the research.

If you have questions regarding this study you may contact Dr. Betty Ashbaker at

422-3857. If you have questions regarding your rights as a participant in research projects, you may contact Dr. Renea Beckstrand, Chair of the Institutional Review Board for Human Subjects, 422 SWKT, Brigham Young University, Provo, UT 84602; phone, (801) 422-3873; email, renea_beckstrand@byu.edu

Appendix C

Samples of Different Types of Questions

Multiple-choice.

1. (Multiple-choice). Which of the following would be considered a responsibility of a paraprofessional?

    a. Report progress to parents

    b. Design lesson plans

    c. Create behavior plans

    d. Supervise students

Matching.

2. Match the following terms with the corresponding field.

    e. Paralegal _____health field

    f. Paraprofessional _____law

    g. Paramedic _____works with teachers

    h. Paraeducators _____works with related servers

True-false.

3. Choose true or false and defend your answer.

    Students are naturally drawn into correct behavior patterns if their instruction is effective.

    Answer: True. Accept such reasons as:

    Effective instruction not only ensures learning, it ensures engagement, which is a critical requirement for learning to take place and a natural controller of behavior

Fill-in-the-blank.

4. Fill-in-the-blank.

   _____ is one of the most pro-active measures you can take towards managing behaviors in the classroom.

   Answer: Effective instruction

   Short answer.

5. The text lists four ways in which we communicate, intentionally or otherwise. Other than using words, pick one type and explain how it is a communication system.

   Essay.

6. (Essay) According to the chapter, there are definite differences between paraprofessionals and teachers. What are they? In your opinion, state why they are important.

Appendix D

Sample Evaluation Question with Instructions

Read the section text on the first page carefully. After you have finished reading it, turn the page completely over and put it under your booklet. You cannot refer back to the first page after turning it. Read the question and mark or write an answer. Then, rate the question according to its difficulty. After you have finished answering all the questions, turn the page and move onto the next section of text.

Question: You are a veteran paraprofessional and a new paraprofessional will be working in the same classroom as you. What suggestions can you give to this individual to help him or her obtain training?

Rate this question (circle your choice).

Easy Average Difficult