DOCUMENT RESUME

ED 044 065 FL 001 973

AUTHOR	Pimsleur, Paul
TITLE	New Approaches to Old Problems through Testing.
SPONS AGENCY	Wisconsin State Univ., Whitewater.
PUB DATE	70
NOTE	8p.; Speech presented at the Sixth Annual Whitewater Foreign Language Conference, Whitewater, Wis., Mar 1970. Reprinted from the Canadian Modern Language Review; v27 n1 p23-32 Oct 1970
JOURNAL CIT	Wisconsin Association of Foreign Language Teachers Bulletin; v56 n1 p2-9 Fall 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.50
DESCRIPTORS	Achievement Tests, Aptitude Tests, Dropout Problems, *Equivalency Tests, Instructional Program Divisions, Item Analysis, *Language Instruction, *Modern Languages, *Motivation, Student Attitudes, Test Construction, Test Interpretation, Test Reliability, *Tests, Test Validity

ABSTRACT

Positive values inherent in well-planned tests used in language programs are discussed in this article. The author argues that the problems of dropout rate, underachievement, discipline, and negative student attitudes can be alleviated through testing programs which include: (1) an aptitude test for selecting and sectioning students and for guidance of those with foreign language learning problems; (2) brief but frequent quizzes for learning purposes; (3) short progress tests for learning and motivation; (4) periodic review tests for motivation as well as grading; (5) reliable mid-term and final examinations strictly for grading; (6) an oral examination before a committee once or twice every academic year; and (7) a standardized proficiency test every second or third year for control of the program and for placement. Discussion of progress tests, test validity, item analysis, articulation, placement, and aptitude tests focuses on the need to develop complete and sound testing programs. (Author/RL)

The WAFLT Bulletin, Fall 1970

NEW APPROACHES TO OLD PROBLEMS THROUGH TESTING*

Paul Pimsleur
State U. of New York at Albany

Let's name the most persistent problems we can think of -- the ones that plague foreign language teachers and never seem to go away.

DROPOUTS. Nationwide, half our students drop the language after one year, nine-tenths after two.

UNDER-ACHIEVEMENT. We try hard to perfect our teaching methods, and yet there are still some students, too many for comfort, who don't grasp the subject at all.

DISCIPLINE PROBLEMS. Bored by the language lesson, some students become restive and disrupt the class.

NEGATIVE ATTITUDES. Is your experience like mine: when I meet a new person and he finds out I'm a foreign language teacher, there's a good chance he'll tell me, unsolicited, what a bad time he had with French or Spanish in high school. Foreign languages share with mathematics the reputation of least-favorite, most-feared school subject. Many people carry this attitude into their adult lives and later infect their children with it.

I'd like to take a fresh look at these problems with you today, and see whether they can be approached in a new way -- through testing.

Many people are misinformed about tests. They think that tests are only for giving grades, forgetting that, to most students, grades are a punishment. The only good grades are A and B and most students don't get them, so the majority are punished by their grade. Yet Skinner assures us that in the long run reward will induce more learning than punishment. Punishment causes people to exert only the minimum effort necessary to avoid it; to work just hard enough to keep from failing, but not hard enough to really learn the language. Grading is important, of course, and I'll come back to it before I'm finished. But first let's talk about a testing program that will show students how much they have learned and how much they can accomplish, not how much they have failed to learn.

First of all, tests help students to learn. To illustrate, I'd like those of you who know French to join me now in a typical grammar drill. This drill is on the compound past tense, the passé composé. I'm going to say a sentence in the present tense, and you are to say the same thing in the past. For example if I say "Je parle au téléphone", you would say "J'ai parlé au téléphone." Let's begin.

Stimulus:	Confirmation:
Je parle au téléphone.	J'ai parlé au téléphone.
Robert pose une question.	Robert a posé une question.
Marie donne la réponse.	Marie a donné la réponse.

Good. And now, to illustrate how close this drill is to being a test, I'd like you to give yourself a check-mark for every time you get one right, and an X if you get it wrong. Keep your score as we continue the same drill. They'll get harder now.

Stimulus:	Confirmation:
Marie donne la réponse.	Marie a donné la réponse.
Je reçois le paquet.	J'ai reçu le paquet.
Je l'ouvre.	Je l'ai ouvert.
Ma petite soeur me regarde.	Ma petite soeur m'a regardé.
Mais elle ne voit rien.	Mais elle n'a rien vu.

Now you've scored 5 items; add up the check-marks. How many got all five right? Who got four right?

The question is: was this an exercise or a test? I don't know. Call it perhaps a "learning test." What is important is the frame of mind it put you in. You were eager to learn. Drills like that often are tedious, but this one wasn't. Why? Well, it seems there's something about human beings that responds to a challenge. You were eager to try to answer my questions; you knew you would find out right away whether you had given the right answer, and that if you were wrong, you'd learn what the right answer should have been, so you'd be more likely to give it next time. That state of eagerness and receptivity is exactly the frame of mind we want our students to be in.

So tests help students to learn. Secondly they also help teachers to teach.

A teacher puts forth a lot of effort trying to get a grammar point across; you work hard to keep the drills lively and effective. As you're doing them, you watch your students. Do many hands go up or

*This address was delivered at the Sixth Annual Whitewater Foreign Language Conference in March, 1970. It is reprinted here with the permission of the Editor from the Canadian Modern Language Review.

few, do the faces look eager or are they puzzled, do the answers come quickly or are they painfully slow; these are your clues to whether you're getting your point across. Unfortunately, they're not reliable clues. All too often you find out, when you give a quiz, that those eager faces fooled you and that many students didn't grasp the point at all. You need insurance against these surprises, and your best insurance is to give brief quizzes often. Actually, there should be a quiz after every grammar point; one that is short and easy to correct. Such quizzes keep you aware of what has been learned and what hasn't, so you can take immediate steps to keep students from falling behind. A quiz should pinpoint the problems. For instance, in the drill we did a few minutes ago where you scored yourselves: raise your hand if you got the fifth item wrong. That one was Mais elle ne voit rien, to which you were to say Mais elle n'a rien vu. The problem there was the placement of rien before the past participle; I see that some of you got it wrong, so I know I have to drill this point again.

So tests help students to learn, and they help teachers to teach. Thirdly, they help to maintain motivation.

Everyone who has learned another language knows that it is a long and often discouraging process. If the students really knew how far they have to go to master it, few of them would have the courage to undertake it. To keep their interest high and prevent their getting discouraged, we've got to break this long journey down for them into manageable steps, punctuated by periodic tests. The right test given at the right time proves to the student he's making progress; convinces him he can say and write things in French or Spanish this week that he couldn't say or write last week. A "progress test", given once a week or so, can help to do this.

Similarly, when you're walking down a long road, it is encouraging to glance back over your shoulder once in a while to see how far you've come. That's what a review test does. Given every four to six weeks, it marks the plateaus that the student has reached. It also gives the teacher a grade he can use for the 6-week marking period.

So far I've mentioned: learning tests, progress tests, review tests. They all aim to reward the students, not to punish them; to convince them that they have made progress. That's why none of these tests aim at a "normal distribution" of scores; there is no point in telling half the class that they are "below average" when in fact they may have made real progress in learning the language. Only the few students who have not made even minimal progress should be told they are "below par" -- and then they should be helped, urged, and practically forced to climb back up, not allowed to fall farther and farther behind. I contend that the vast majority of students ought to be making enough progress so that they merit a decent grade; if they are not, then the language program itself -- the textbook, the methods, the course objectives, and the syllabus -- need looking into.

A study was done a few years ago which proved that foreign languages are more 'sequential' than any other subject in the curriculum. In other words, next year's grades depend on this year's grades; so much so that a student is actually more likely to do worse next year than he is to get the same grade again. Evidence shows that a student who gets a 'C' in Spanish One will probably get a 'D' or an 'F' in Spanish Two. I think a strong case can be made for considering 'A' and 'B' as the only acceptable grades for promotion in foreign languages; any lower grade means the student knows too little to do the work at the next level.

Learning tests, progress tests, review tests. In these days of audio-lingual emphasis, we naturally test speaking and listening as well as reading and writing. But speaking is hard to test. It must be done one student at a time. (I'm ignoring for the moment the small number of schools that have a full-record lab and manage to keep it in working order.) So the best the teacher can do is to score a brief sample of each student's speech from time to time, and this is not enough to put teeth into the pursuit of the speaking objective. It is axiomatic that students study for what they are graded on. If we want them to take the speaking skill seriously, we've got to grade them seriously on it. The tests we've talked about so far do test speaking, among other things; but they don't go far enough. They do not test whether a student can engage in free conversation within reasonable limits. This is, after all, the objective we're trying to reach, and we have no hope of reaching it unless we periodically test for it.

We must borrow an idea from European education: the oral examination. The students appear, one at a time, before a committee of two or more language teachers. Before coming in, they've had three minutes to prepare a topic pulled out of a hat; something they're familiar with, like a story the class has read. They can jot down a few notes to help them remember what they want to say, but not whole sentences. Then they come before the oral testing committee. They tell their story, and one of the teachers prompts them when they hesitate. After the story, one of the teachers converses with them about everyday life: the language class, the school, the student's house, family, or friends. The entire test takes no more than three or four minutes per student, which is enough time to find out how well he can speak. He is graded separately for accent, fluency, and vocabulary-syntax.

Administrative problems? Yes, an oral exam requires some shifting of schedules; putting two classes together, for example, so the two teachers can form a committee. That's why I only suggest doing it twice

a year, at the midterm and the final. But think of what you achieve. An oral exam is a powerful motivator. Even one such exam a year may be enough to keep the students practicing their speaking all year long, since it cannot be crammed for, and only a continuous effort, a year of speaking practice, will prepare them. The oral exam creates a mood where students want to speak the foreign language among themselves, to get the practice. It can light a fire under a tired audio-lingual program.

An ideal testing program is taking shape: teaching and learning tests almost daily; progress tests once a week; review tests every six weeks; an oral exam once or twice a year.

Now what about grades? Most of the tests we've mentioned so far will be on the easy side, to keep the majority of students moving along with the program. But we've still got to give midterm and final grades, and to do this we make up midterm and final exams. I'd like to show you how these teacher-made exams can be improved; how you can make sure they are doing a fair and adequate job of measuring your students' achievement.

Teachers usually make up an exam, give it, mark it, and forget it. Did you ever wonder whether your exam was fair; not too hard, too easy, or too tricky? Most teachers do; I've found that hardly any teachers are satisfied with their tests.

To get closer to the problem, let's do a little experiment. I'm going to show you a passage and some questions about it from a Spanish reading test. I'll ask those who know Spanish to state how they think students are likely to do on these questions. Here is the passage.

Mozo. - ¿Qué desean tomar?

Lola. - Quiero el pescado del día, arroz, ensalada y café.

Rosa. - Yo quiero lo mismo que Lola, y además la cuenta.

Now I'll show the questions, and I'd like you to estimate what percentage of students would answer correctly at the end of one year of high school Spanish. Here is the first item.

1. Las muchachas están en	Difficulty	Discrimination
1. una sala grande.	93*	.34*
2. una cocina.		
3. un restaurante.		
4. una playa.		

How many think about 25% would get it right? About 50%? About 75%?

This test was given to 244 students. Question one was answered correctly by 93% of them, so it is a very easy item. The other figure, .34, indicates its discrimination power. It means this question distinguished fairly well between good and poor Spanish students; the few students who didn't get it right were poor students, not good ones.

Try the next question.

2. Probablemente ¿qué comen?	Difficulty	Discrimination
1. el desayuno.	39	.08
2. la comida.		
3. el almuerzo.		
4. el bocadillo.		

What do you think the figures are? Did 25% get it right? About 50%? 75% or more? And which ones got it right; did the item distinguish well or poorly between students who have learned their Spanish and those who haven't?

This item was answered correctly by 39% of the students, so it is fairly difficult. But the important thing is that figure of .08 under discrimination; that means it is discriminating very poorly. It is a very bad item because the good students are missing it just as often as the poor ones. The students are equally divided between choice 2: "la comida", and choice 3: "el almuerzo". This item is badly in need of revision.

*In oral presentation, these figures were shown after the audience guessed at them.

What do you think of the 3rd item?

3. ¿Quién quiere pagar?	Difficulty	Discrimination
1. un mozo.	53	.54
2. las dos.		
3. Lola.		
4. Rosa.		

Was it hard, easy, or about right? Well, it was about right, about half the students answered it correctly, and it discriminated well.

I hope my point is clear. It is very difficult for a teacher to judge whether questions are just right for a given group of students. No matter how experienced he is, he will misjudge how hard a question is, and he will sometimes write questions that trick people instead of testing them -- like the "almuerzo" and "comida" question. You've all had the experience of making up a question where you've had an idea in mind that seemed very clear to you, and then your students have seen something quite different in it than you intended. To guard against errors of this sort, a test has got to be pretested and then item analyzed, as we just did for those Spanish questions. Naturally, you can't give it as a pre-test to the same students who'll then take the revised test. What you must do is item analyze the next major test you give (use the criteria of difficulty and discriminatory power as we just did; any book on educational measurement will explain how). Then revise the questions to remove the flaws you find and put the questions into a card file, one question to a card, with the data from your analysis on it. By doing this regularly, you will build up a file of test questions from which to draw for future tests.

Let's take a Listening Comprehension item as an example. Suppose you had made up an item to test certain features of the verb courir, 'to run'. Here is how the item might look.

ITEM FILE, front

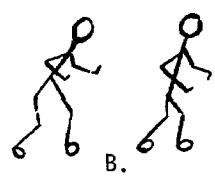
French Listening Test

Text
Unit IV

Type: Multiple-choice
Content: le/les distinction
Vocabulary: garcon, court

(Le garcon court.)

A.



B.

C.

A.

You want to list the skill it is testing (Listening), the name of the textbook and the chapter or unit where it occurs, the type of item (multiple-choice), the item content (choices A and B test the le/les distinction), and the vocabulary employed. Then after you have given the item in a test, you enter on the bottom of the card or on the back the significant information about how it worked.

ITEM FILE, back

Date	No. of papers	Difficulty	Discrimination
Nov. 13, 1968	28	75%	7/8 - 4/8 = 38%
Dec. 7, 1968	28	89%	8/8 - 6/8 = 25%

The date when it was given, the number of students who took it, and then the two features we saw a few minutes ago, difficulty and discrimination. The difficulty index is simply the percent who got it right. The index of discrimination is obtained by calculating how many more good students got it right than poor students. You do this by taking the top 30% and the bottom 30% of the class. In this example, the first time the item was given, seven out of eight good students got it right versus only four out of eight poor students. This works out to a 38% difference. It sounds a little complicated, perhaps, but in practice it is very simple to do. In fact, you can have the class do it by a show of hands as to how many got each item right, with one student doing the calculation at the board. It is simple enough to do entirely in the foreign language and provides excellent practice in the realistic use of numbers.

Returning to our main point, we've now added to our list of tests a midterm and a final, made up by a committee of teachers from a file of pre-tested items and administered school-wide or district-wide to insure articulation in the foreign language program.

That word articulation brings me to an especially pesky problem: placement. In this day of multiple tracks and levels, students arrive in high school with backgrounds that can vary from no foreign language study to eight years of study (or at least of exposure). I remember one boy I talked to after a Spanish class. He had acted very bored during class so I made it a point to ask him why he didn't like Spanish. "Oh, I like it all right," he said. "But this is the fourth time I'm taking Spanish I." He had a radio course in the 3rd grade, a television course in the 5th, a resource teacher in the 7th grade, and here he was in the 9th grade starting Spanish over again because the school had no procedure for placing him where he belonged -- either in Spanish II or in a special class for students who already knew a little Spanish.

In this era of mass education we are all concerned lest the individual child, with his particular knowledge and abilities, be sumberged in the system and his special needs forgotten. Our placement procedures must of course assign the mass of students to the classes they belong in, but they should also recognize when an individual student needs special guidance and bring his case to our attention.

A high school enrolls students funnelled in from several junior highs, plus some whose families have just moved into the district. In some districts, economic conditions are so unstable that a large percentage of school seats are filled with different students at the end of the year than at the beginning. Some seats may have changed occupants two or three times.

To cope with the problem of placement, you must begin by rejecting easy solutions that don't work. Two years of language in high school do not necessarily equal one year in college, though this is the credit commonly given. Similarly, two years of language in junior high do not equal one year in high school. A year's work in one high school is rarely the equivalent of a year's work in another, since the specific points of grammar and vocabulary a student learns from one teacher are never quite the ones his next teacher expects him to know. This is so even in the same school, and even more true in different schools, with their different books and methods. Never forget that languages are the most sequential subjects in the curriculum, so articulation is of capital importance.

If you want to place incoming students properly, you've got to have a hard and fast measure of your present students' achievement as a standard of comparison. Teacher-made tests won't do, because they change every year as your program evolves. You want a test that remains constant over the years.

You also want an unbiased external measure of your program. Not a test that perpetuates the idiosyncracies of your particular teachers, textbooks, and methods, but one that measures how far along students have come toward language mastery, irrespective of the methods used. In short, not a test of achievement but of proficiency. Moreover, you want a proficiency test that has been standardized, because this guarantees you that it is reliable, that it will continue to be available for many years, that you'll have the convenience of machine-scoring and tabulation, and that you can compare your student's proficiency against the proficiency reached elsewhere.

There are two sets of standardized tests on the market at present.

MLA-Cooperative Foreign Language Tests. Educational Testing Service, Princeton, New Jersey.

Pimsleur Foreign Language Proficiency Tests. Harcourt, Brace & World, Inc., 757 Third Avenue, New York 10017.

Both of these sets offer the range of information and services you would expect from a commercial publisher. The price per student is quite reasonable, especially if you do your own scoring, because much of the test material can be re-used. There are separate tests of the four skills, with tapes for measuring listening and speaking. I suggest you write to the publishers for gratis specimen copies.

Our question, you remember, was how to place incoming students, with their diverse backgrounds. We start by giving a standardized test battery to every one of our language students, at all levels, at the end of the year. (Note that we'll only have to do this every second or third year to keep our figures up to date.) We tabulate their average scores, which may look like this for the Listening skill, for example.

	Listening
First year	50
Second year	61
Third year	72
Fourth year	84

We add the standard deviation, to give us an idea of the range of scores in each group.

	Listening
First year	46 - 50 - 54
Second year	55 - 61 - 66
Third year	66 - 72 - 78
Fourth year	78 - 84 - 90

These ranges mean that 2/3rds of the students got scores within these limits. For instance, in the top line the average score achieved by first-year students was 50, and the majority of them got scores between 46 and 54.

We do the same thing for the other skills. Here we've added Reading scores, for example.

	Listening	Reading
First year	46 - 50 - 54	51 - 55 - 59
Second year	55 - 61 - 66	61 - 67 - 73
Third year	66 - 72 - 78	72 - 76 - 80
Fourth year	78 - 84 - 90	78 - 83 - 88

Finally, we make a composite table from the scores in all the skills. (Note that for practical reasons many schools find it difficult to give the Speaking Test and only use the other three skills.) We now have a set of local norms.

L O C A L N O R M S

	Four Skills
First year	48 - 52 - 56
Second year	59 - 64 - 69
Third year	69 - 74 - 79
Fourth year	79 - 83 - 87

Next autumn, the first thing we'll do is give the incoming students who need placement the same test, and place them by the results. Suppose a student gets a composite score of 52? According to our table he's only as proficient as one of our average or "C" students at the end of one year. He probably does not know enough to do second year work in our school and should start again. Better yet, if there are enough students like him to warrant it, we may make a special section for them. Another incoming student gets a score of 57; he is superior to most of our first-year students, and is probably prepared for second-year work. Often, we may look at the particular skills, too, to see that the students have at least a minimal command of listening comprehension, speaking, reading, and writing. If they are coming to us without one or another of these skills, we can find it out this way and gear our teaching to make up for the deficiency. My purpose is not to tell you how you must use the information from standardized testing, but to point out that you can pursue your educational aims much more effectively if you have such information than if you don't.

Besides, you receive many fringe benefits from giving a standardized test every couple of years. You can see whether all four skills are progressing smoothly, or if listening, say, is being overlooked. You can see if the new lab you've installed, or the new textbook you've selected, has made a difference in your students' proficiency. A standardized test, given every 2 or 3 years, provides objective evidence of the progress you've made and the progress you still need to make in improving your foreign language program.

Our list of tests is almost complete, but there is one important problem we haven't talked about; placement of beginning students. In most schools a decision is made in the springtime as to which students

are going to begin foreign language study next year. Typically, this decision is made in the 6th grade about beginning in the 7th, or in the 7th about beginning in the 8th. Unfortunately, it is made in ways that are demonstrably unfair to a sizeable number of students. Many schools don't even attempt to bring any evidence to bear; they just ask the homeroom teacher to make the selection and she bases her decision on her subjective evaluation of the student's "readiness," whatever that is. Other schools use English grades or IQ to help them decide who's to take a foreign language. A few years ago, I did a study of the accuracy of various commonly-used methods, and here is what I found:

Predictor	Correlation with FL grades
IQ	.46
English grades	.57
Grade-point average	.62
Aptitude Battery	.62
GPA and Battery	.72

The decimal numbers on the right are correlation coefficients. They show how closely the particular predictor relates to FL grades. A perfect relationship would be 1.00, but of course we never reach that. The first line says, for example, that IQ relates .46 to foreign language grades; this is a modest correlation.

As you can see, IQ and English grades do not correlate as highly with language grades as grade-point average does. So schools should at least use grade-point average, which they have readily available.

In this experiment, I also gave a one-period aptitude battery and it predicted grades as accurately as grade-point average. The correlation is the same, .62, yet the two measures are quite different. Grade-point average reflects overall school achievement, whereas the battery measures the specific skills involved in learning a foreign language. Suppose you put the two together, wouldn't they complement each other? Yes they do; they raise the correlation to .72, which is a terrific increase considering that the higher the correlation already is, the harder it is to improve it further.

In examining these correlations, let's not lose sight of their purpose: to guide students correctly. Behind the correlations are children, and the higher the correlation the fewer the children who will be misplaced. The fewer the errors of assigning a student to a class where he won't do well, or barring a student from a class where he would have done well. All schools make decisions about who is to take language in 7th grade and who must wait until 8th or 9th; who is to take French, who Spanish, and who German. What I am arguing is that these decisions have got to be made accurately; they affect the lives of the students and success or failure of the school's language program.

Did you know that there are under-achievers in every language class? I mean students who have more trouble with languages than with their other subjects; maybe they get B's elsewhere and C in Spanish, or C's elsewhere and they fail it. I've looked into this problem in junior high schools, high schools, and colleges, and I can tell you that 10 to 20 percent of the students in any language class you care to choose are getting a significantly lower grade in language than in their other subjects. In a class of 30, that means 4 or 5 students. The teacher usually mistakes them for poor learners because he hasn't gone into their school record. But they are under-achievers, which isn't the same thing. They study conscientiously (many under-achievers put in inordinate amounts of time on language homework to the detriment of other subjects), yet they can't seem to get their grades up to the level everyone expects of them from their past performance. So they and their parents are baffled and frustrated.

An aptitude test, of which there are two in common use,

Modern Language Aptitude Test. The Psychological Corporation, 304 East 45th Street, New York 10017.

Pimsleur Language Aptitude Battery. Harcourt, Brace & World, Inc., 757 Third Avenue, New York 10017.

helps diagnose these problems in advance so the students can be helped before he falls behind the class. We all know about the terrible dropout problem in foreign languages; 50% of the students quit after the first year. Surely no one is more ready to drop language than a student who can't keep pace with his class, and I believe that one way to prevent dropouts is to make sure students are placed in the right class to begin with, and that special learning problems are diagnosed early.

Now our ideal testing program is complete. I've recommended: an aptitude test for selecting and sectioning students and for guidance of those with foreign language learning problems; brief but frequent quizzes for learning purposes; short progress tests for learning and motivation; periodic review tests for motivation as well as grading; reliable mid-term and final examination strictly for grading; an oral examination before a committee once or twice a year to keep the oral goal alive; and a standardized proficiency test every second or third year for control of the program and for placement.

I D E A L T E S T I N G P R O G R A M

Type of Test	When Given?	Purpose?
Aptitude test	previous year	sectioning, guidance
Teaching/learning quizzes	almost daily	learning
Progress tests	weekly	learning, motivation
Review tests	every 4-6 weeks	motivation, grading
Midterm and final exams	mid-year; end-of-year	grading
Oral exam	1 or 2 per year	grade and motivate speaking skill
Standardized test	every 2-3 years	program control, placement

Most schools do some of this, but hardly any do it all. Yet I think you would find that some of your most persistent problems -- dropouts, student apathy, under-achievement, hard-to-control classes, negative attitudes toward language -- can be resolved or at least reduced by a judicious use of tests.