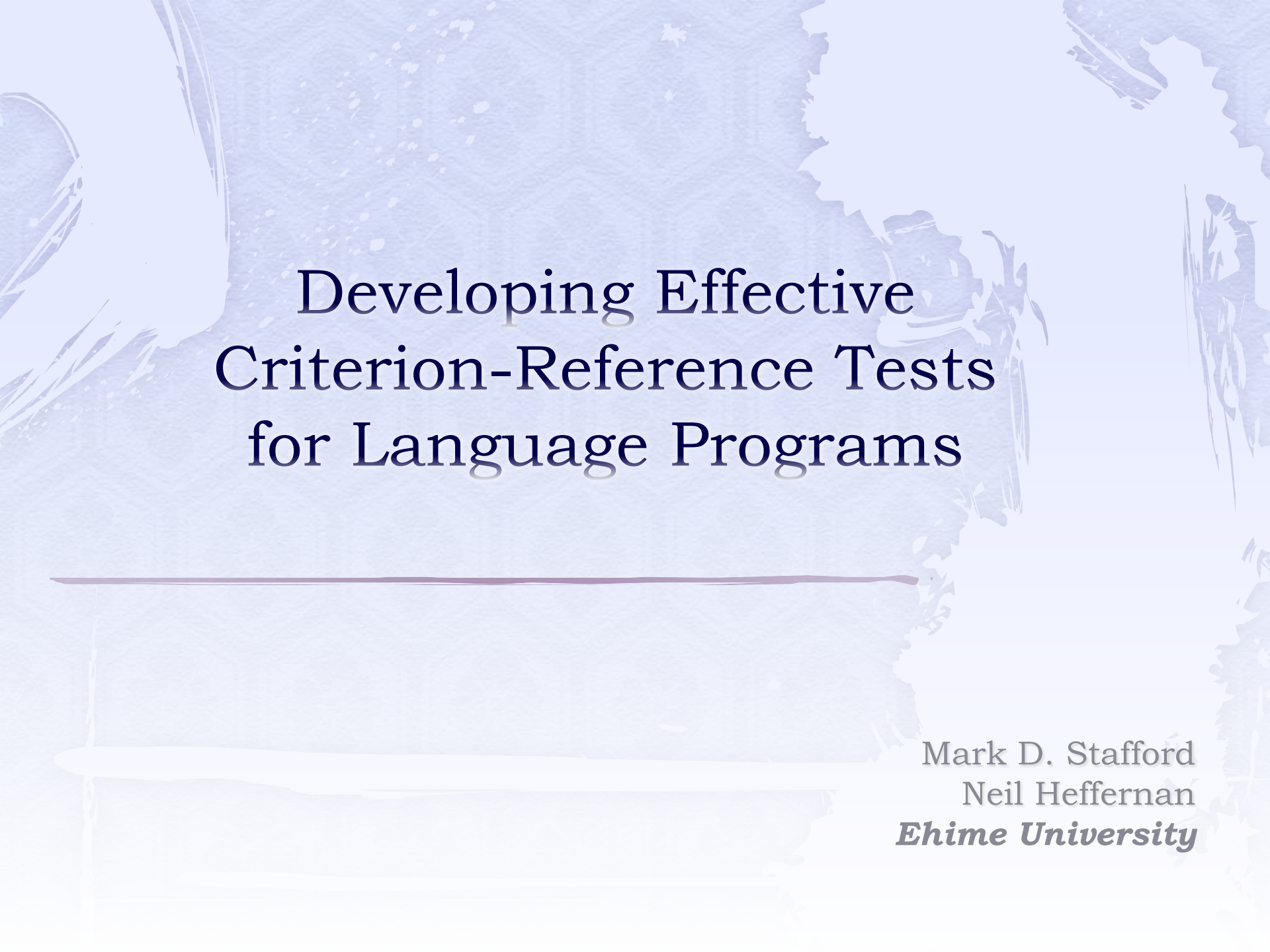


Developing Effective Criterion-Reference Tests for Language Programs

Neil Heffernan and Mark D. Sttafford

Reference data:

Heffernan, Neil and Sttafford, Mark D. (2009) Developing Effective Criterion-Reference Test for Language Programs. In: Reinelt, R. (Ed.) (2009) Into the Next Decade with (2nd) FL Teaching Rudolf Reinelt Research Laboratory EU Matsuyama, Japan, p. 294-308.



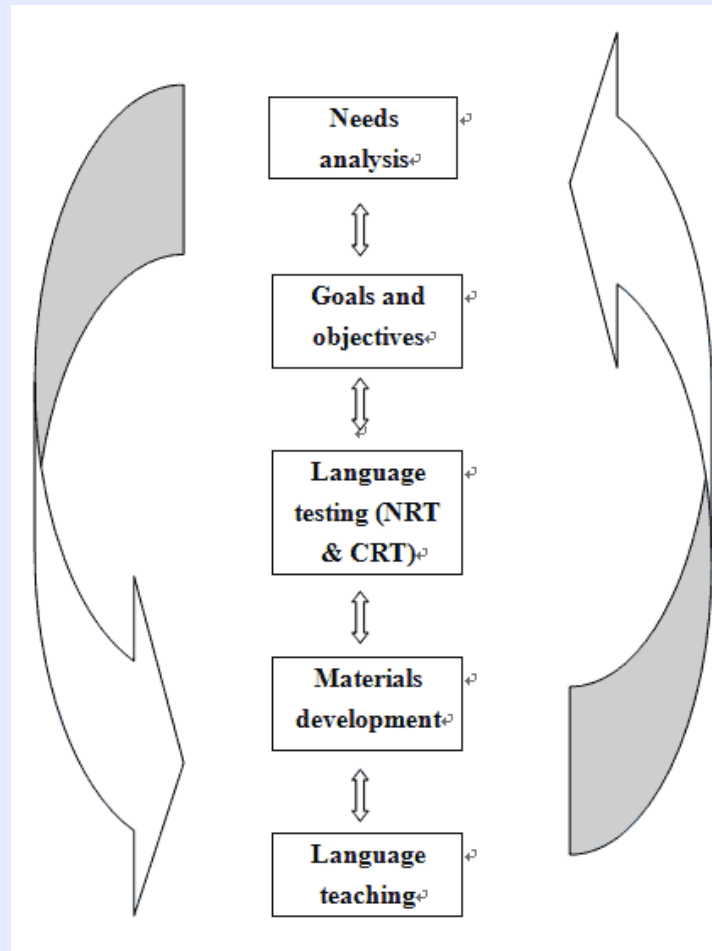
Developing Effective Criterion-Reference Tests for Language Programs

Mark D. Stafford
Neil Heffernan
Ehime University

Introduction

- ◆ The English Education Center (EEC) at Ehime University
- ◆ Curriculum development:
 - four skill-based classes for 1st-year students
 - one textbook created in-house for each skill

Design of language curriculum



Criterion vs. Norm-referenced tests

- ◆ **Criterion-referenced:** designed to measure well-defined and specific objectives: particular to a particular course or program. Each student's score is meaningful without reference to other students' scores. Scores indicate students' knowledge or skill in the objectives learned.
- ◆ **Norm-referenced:** designed to measure global language abilities. Each student's score is interpreted relative to the scores of all other students who took the same test.

see Brown (1996)

Global Test of English Communication (GTEC)

Currently used twice a year (June and December)
as a measure of students' abilities

Created by Bennese Corporation, a Japanese
company

Made for Japanese students with Japanese
instructions on the test

Consists of reading and listening sections only
(writing has recently been taken off the test)

EEC: Goals and “Can-Do” lists

- ◆ The current curriculum is based on “Can-Do” lists that were compiled three years ago.
- ◆ One “Can-Do” list for each skill. For example:

Listening

Can understand information about familiar topics related to everyday life
Can understand an introduction and explanation of a school or company

Speaking

Can order while looking at a menu
Can express desires and preferences to a clerk while shopping

Reading

Can understand articles about topics of interest
Can understand explanatory texts written for the general public

Writing

Can write simple texts introducing where one lives
Can write summaries of content heard or read about matters of interest

GTEC: Problems

- ◆ A norm-referenced test for a program that needs a criterion-referenced test.
- ◆ Counts for 30% of students' mark for two classes each semester.
- ◆ Not a reliable indicator of student abilities.
- ◆ Not a reliable indicator of what goes on in the EEC.

Test application and development

Listening test

- based on the textbook and curriculum of the listening course in the EEC
- created with the specific goals of the listening program in mind
- designed to measure the amount of learning students accomplished on each objective
- students knew the general content of the test because they had covered it for 15 weeks in class

Examples of test questions

Good example:

16&18 refer to a conversation about housing.

16. Who is the woman?

80% A. A customer.

10% B. A housemate.

7% C. A real estate agent.

3% D. A landlord.

Too easy:

46&47 refer to a conversation at a hotel.

46. What is the problem with the room?

1% A. It's far.

1% B. It's expensive.

2% C. It's small.

95% D. It's noisy.

Too hard:

48-50 refer to a talk about registering for classes.

49. What should someone do after clicking the "accept" button?

50% A. Print the page.

20% B. Make a reservation.

8% C. Interview a professor.

23% D. Select classes.

Inverted:

6-8 refer to a talk at a school.

6. What can students do at the administration building?

67% A. Check their work schedule.

6% B. Find the health center.

23% C. Check class enrollment.

4% D. Visit lecture halls.

Results

- ◆ Some students performed well on the test: 60-90% success rate
- ◆ Since the purpose of a CRT is for students to do well on the test, this test was considered successful
- ◆ The purpose was for students to pass the test, not to rank them. "Perfect" on this test would be for everyone to get 100%.

Reasons to improve

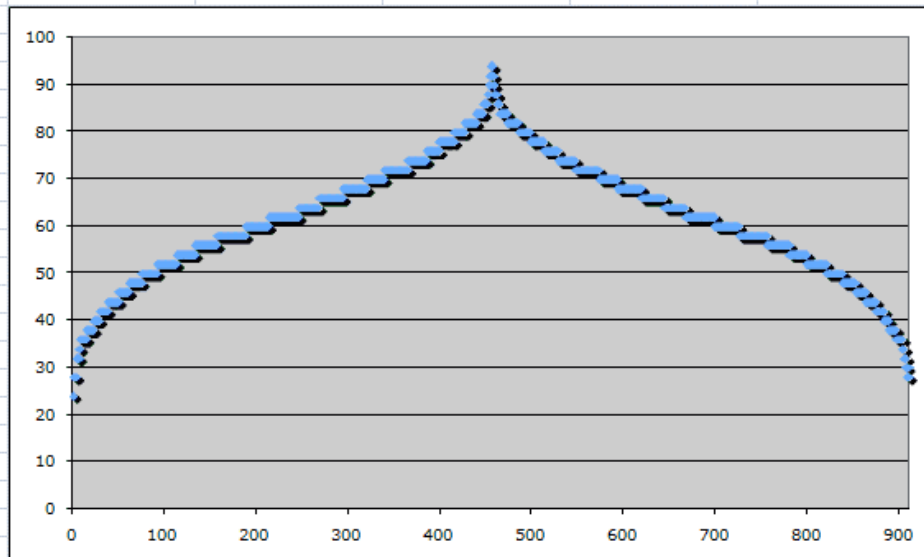
As demonstrated, some items were:

1. too easy (can answer without learning the material)
2. level appropriate
3. too complicated

Also, the recording was done in-house, so the quality could be improved

Results by faculty

	All	Medical	Education	Night	Law & Letters	Agriculture	Science	Engineering	FB
N	912	106	133	81	71	44	116	329	32
MEAN	62	70	67	66	65	61	60	57	51
MODE	62	72	58	78	64	66	62	58	50
MEDIAN	62	72	68	66	66	64	61	58	50
MIDPOINT	59	59	60	67	64	56	54	58	52
MAX	94	94	88	90	92	84	84	88	72
MIN	24	24	32	44	36	28	24	28	32
RANGE	70	70	56	46	56	56	60	60	40
STDEV	13	12	12	10	11	14	12	12	11



Future directions

- ◆ Improve items on the listening test
- ◆ Record scripts professionally
- ◆ Currently, we are developing criterion-referenced tests for reading, writing and speaking

Thank you!
Questions?
