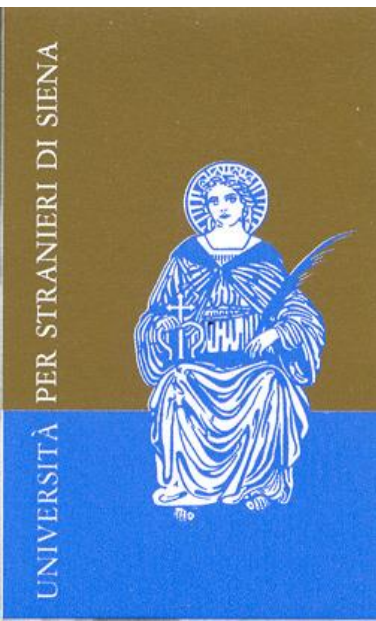

Language assessment for migration and social integration: a case study



Paola Masillo

Università per Stranieri di Siena

Scope of the study

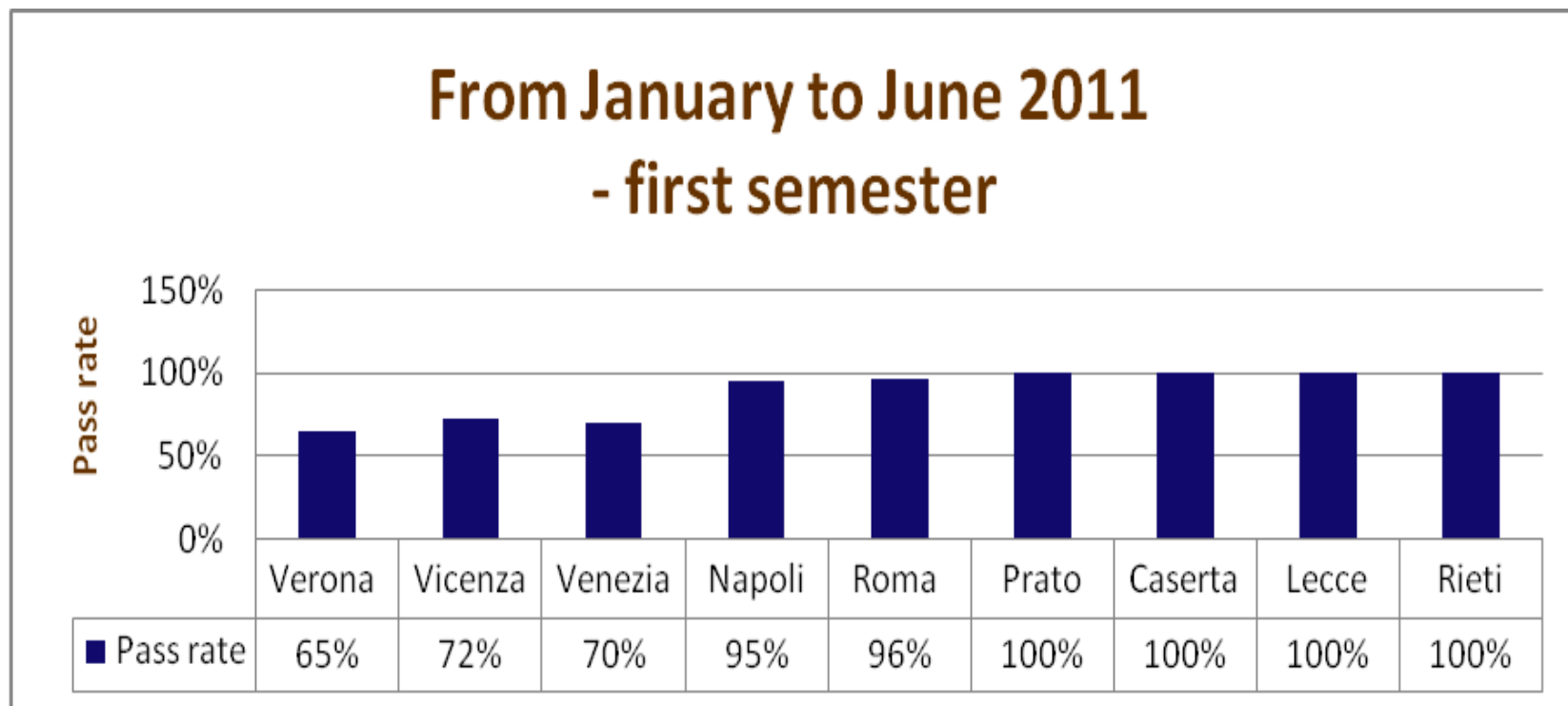
- Partial results of a PhD research project dealt with a *comparability study of two Italian language proficiency tests delivered in Italy* in order to assess language competence of migrants.

Legislative decree 286/1998, art.9, *Immigration Act* – as modified by law 94/2009; **Ministerial decree June 4th 2010** on *Methods of implementation of the Italian language test*



Scope of the study

Table I



Source:

Ministry of the Interior - Department for Civil Liberties and Immigration



Context of research

I. Increasing mobility and migration and easier international and intra-national communication:

→ direct implications on language uses and on language policy (Council of European Union, 2002; European Commission, 2008; Blommaert, Leppänen & Spotti, 2012)

Not all kinds of multilingualism are considered as having the same value: “not all forms of multilingualism are productive, empowering and nice to contemplate” (Blommaert, Leppänen & Spotti, 2012: 1).



Context of research

- II. “Shift in the understanding of the functions, status and roles of language tests. From tools used to measure language knowledge, they are viewed today more and more as instruments connected to and embedded in political, social and educational contexts” (Shohamy, 2007: 117).



Context of research

III. Publication in 2001 of the CEFR: the most important reference document and operational tool in the fields of language learning, teaching and assessment (CoE, 2001).

- Language tests are shielded by policies based on the ideological premise that success in language tests not only shows *a clear willingness to integrate* on the part of the person taking the test, but also provides *the key to success in the workplace and in society as a whole*.
- The power of tests becomes even stronger when test criteria such as rating scales affect language policy and are used to make decisions about people's lives (Shohamy, 2006).



Overview of the study

- Legislative decree 286/1998; Ministerial decree June 4th 2010:
 - **Italian language proficiency test for migrants** asking for EC residence permit of long-term.
 - **Proficiency level: CEFR A2**
- *VADEMECUM, Indicazioni tecnico-operative per la definizione dei contenuti delle prove che compongono il test, criteri di assegnazione del punteggio e durata del test (MIUR, 2010):*
 - **Test design:** Listening, Reading and Written Interaction
 - **Assessment criteria**



Written interaction test

- **One task**

Table 2

CEFR Level A2	
Correspondence	Can write very simple personal letters expressing thanks and apology.
Notes, messages and forms	Can take a short, simple message provided he/she can ask for repetition and reformulation. Can write short, simple notes and messages relating to matters in areas of immediate need.

- **10 minutes.**
- The candidate is required **to reply to e-mails, postcards, invitations** or he is required **to fill out a form** (enrollment in courses or schools, personal data, requests for permits, grants, bank accounts, ...).



Assessment criteria

Table 3

Test is performed in a complete and correct way	answers are provided consistently and appropriately to the information required or the form is filled in all its parts (up to 35 points)
Test is performed in a partial way	answers are not always provided consistently and appropriately to the information required or the form is filled in partially (up to 28 points)
Test is not evaluable	answers are not provided or the form is not completed (0 point)



CEFR Level A2

- *Can perform and respond to basic language functions, such as information exchange and requests and express opinions and attitudes in a simple way.*
- *Can socialize simply but effectively using the simplest common expressions and following basic routines.*
- *Can use some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.*

(CoE, 2001)



Research questions

1. Can the imposed assessment criteria describe and measure the language competence of migrants?
2. Can we recognize validity and reliability in measuring and assessing a *basic proficiency level* on the basis of *appropriateness* or *consistency*?



The sample

- **N. 157 test takers**
- **N. 314 scripts → 40 :**
 - 20 scripts for test format 1
 - 20 scripts for test format 2
- **CEFR level: A1⁺ - B1⁻**
- Using the ministerial **35-point scale**
- **11 raters**



The sample

Table 4

Rater n.	Profile	Experience in the field of LT and Assessment (n. years)	CEFR levels of competence assessed so far (mostly)
1	test developer and teacher trainer		11 A1, A2 , B1, B2, C1, C2
2	teacher and rater		4 A1, A2 , B1
3	test developer and teacher trainer		20 A1, A2 , B1, B2, C1, C2
4	teacher and rater		7 A1, A2 , B1, B2, C1
5	rater		1 A2
6	teacher and rater		3 A2 , B1
7	test developer and teacher trainer		18 A1, A2 , B1, B2, C1, C2
8	teacher and rater		4 A1, A2 , B1
9	test developer and teacher trainer		19 A1, A2 , B1, B2, C1, C2
10	test developer and teacher trainer		20 A1, A2 , B1, B2, C1, C2
11	test developer and teacher trainer		18 A1, A2 , B1, B2, C1, C2



Test n. 1

INTERAZIONE SCRITTA



QUESTO È IL MODULO PER LA RICHIESTA DI LAVORO.
DEVI COMPILARE IL MODULO

logo di m0rula www.flickr.com

ALL'AGENZIA DI LAVORO
IL SOTTOSCRITTO - LA SOTTOSCRITTA:

(NOME)

(COGNOME).....

NATO A (CITTÀ DI NASCITA)(STATO)

INDIRIZZO: VIA..... NUMERO

CITTÀ.....

NUMERO DI TELEFONO

PATENTE ITALIANA SÌ [] NO []

CHIEDE DI LAVORARE COME: _____

SPIEGA PERCHÉ

.....
.....
.....
.....
.....
.....
.....
.....

FIRMA _____

DATA ____/____/2013

1) Fill in the form

2) Declare what kind of job you are asking for and explain why



Test n. 2

Send a message to a friend talking about:

- **Job**
- **Market**
- **Study**

SCRITTURA

(10 MINUTI)



Utilizzando da 30 a 40 parole, scrivi un messaggio ad un amico parlando di:

- lavoro
- mercato
- studiare

FINE DELLA PROVA



The method

- **Step I: Familiarization**

“Traditionally, *rater training* attempted to reduce both variability associated with the differences in overall severity and randomness” (Lumley, McNamara, 1995:56).

“Rater training can reduce but by no means eliminate the extent of rater variability. Rater training has the effect of reducing extreme differences – outliers in terms of harshness or leniency are brought into line. [...] Rater training is successful in making raters more self-consistent” (Lumley, McNamara, 1995:57).



The method

- **Step 2: Task design**

Level of clearness	Level of appropriateness	Level of importance for the test taker
Totally unclear	Absolutely inappropriate	Not at all important
Slightly clear	Slightly inappropriate	Slightly important
Clear	Appropriate	Important
Absolutely clear	Absolutely appropriate	Very important



The method

- **Step 3: Rating process**
- **Step 4:** “Can the evaluation criteria imposed by the Handbook of the Ministry of the Education accurately reflect the construct that the test is intended to measure?”

Level of acceptability

Totally unacceptable

Slightly acceptable

Acceptable

Perfectly acceptable



Findings and Discussion

- **Step 2: Task design**

Table 5

<i>Mode</i>			
	Clearness	Appropriateness	Importance
TASK Test 1	3	3	4
TASK Test 2	1	4	2, 3



Findings and Discussion

- **Step 3: Rating process**

Table 6

Example 1, Test 1, Candidate 07	
Chiede di lavorare come	commesa
<i>You ask to work as a</i>	<i>Saleswoman</i>
Spiega perché	perché ho già lavorato in passato come commesa in un negozio di abbigliamento. Per tanto tempo credo di aver la sperienčia per il posto.
<i>Please explain why</i>	<i>because I previously worked as a saleswoman in a clothing store for so long. I think I have the experience for this job.</i>
Example 4, Test 2, Candidate 9	
Mercato	Celine te scrivo questo messaggio x andare il mercoledì al mercato per comprare abigliamenti de questa temporada, anche puoi dire a tua sorella Vilma. Te saluto y aspetto tu messaggio. "Civediamo" verso le 9.00 am. Baci. Maribel
Market	<i>Celine I am writing this message to ask you to go Wednesday to the market to buy clothing for this season. Tell also your sister Vilma. I say good-bye and I wait your message. See you at about 9.00 am am. Kisses. Maribel</i>



Findings and Discussion

- **Step 3: Rating process**

Table 7

		Statistics	
		Writing T1 totalscore	Writing T2 totalscore
N	<i>Valid</i>	20	20
	<i>Missing</i>	0	0
Mean		29.80	29.25
Mode		27	29
Standard deviation		2.707	3.810
Range		10	16
Minimum		24	18
Maximum		34	34

(Green, 2013; Weir, Chan, Nakatsuhara, 2013)



Findings and Discussion

- **Step 4: Validity of assessment criteria**
 - 11 out of 11 raters → **slightly acceptable**
 - **Inter-rater reliability:** 16 out of 55 are good in both cases (.7 or upwards)
- i. *The main difficulty I found was the choice of the appropriate score with a grid of this kind. It seems shortly interpretable and at the end there is the risk of choosing a scoring by chance.*
- ii. *The evaluation criteria of the Ministry of Education seem to me very little detailed and lend themselves to multiple interpretations. An assessment scale from 0 to 28 is very vague.*
- iii. *The evaluation criteria of the Ministry of Education are not adequate, the margins are very wide and leave room for subjective evaluation. It lacks a grid to be followed.*



Final considerations

- Results obtained by both analyses show that while the *inter-reliability* remains quite high, raters were also strongly influenced by *their intuitive impression, expertise and previous experience* (Wu, Ma, 2013).
- This approach suggests that the rating scale *does not address all the elements that influence raters' decisions* (Wu, Ma, 2013) and *does not reflect the criterion and the construct* that the test intends to measure.



Final considerations

- The findings we obtained can be summarized in the following comments, which suggest *the need of a revision of the assessment procedure*, particularly with regards to the *criteria*, in order to enhance *scoring validity* (Wu, Ma, 2013):
 - i. *the message is not completely effective*
 - ii. *I find it very difficult to attribute a score with these parameters, since it does not refer to [...] and efficacy, but only to the completeness of the information.*
 - iii. *I choose this mark because the effectiveness is not good due to a wrong grammar the message is clear*
 - iv. *lack of effective communication*
 - v. *the production is not very effective*



Thank you!

Contact details:
masillo@unistrasi.it

