

# Assessing linguistic competence through self-assessment

**Cristina GALEANDRO<sup>1</sup>, Edo POGLIA<sup>1</sup>, Gé STOKS<sup>2</sup> & Katya TAMAGNI BERNASCONI<sup>3</sup>**

Università della Svizzera Italia<sup>1</sup>, Via G. Buffi 13, CH-6900 Lugano  
Scuola Universitaria Professionale della Svizzera Italia<sup>2</sup>,  
Dipartimento della Formazione e dell'Apprendimento, Piazza  
San Francesco 19, CH-6600 Locarno & Ufficio studi e ricerche<sup>3</sup>,  
Viale S. Franscini 32, Stabile Torretta, CH-6500 Bellinzona  
cristina.galeandro@ti.ch, edo.poglia@lu.unisi.ch, ge.stoks@aspti.ch &  
kathya.tamagni@ti.ch

Nell'ambito del progetto PNR 56 "Ricerca sullo sviluppo delle competenze linguistiche dei giovani ticinesi", è stato chiesto ai partecipanti di autovalutarsi per stabilire i livelli linguistici da loro raggiunti per quanto riguarda le capacità di lettura e di interazione. A questo scopo, i partecipanti hanno fatto capo alla scala delle competenze del Quadro Europeo di Riferimento, indicando il livello raggiunto sulla scala globale (A1, A2, B1, B2, C1, C2) per francese, tedesco ed inglese. In seguito, un campione di un centinaio di partecipanti ha svolto il test Dialang per la lettura. Questo ha reso possibile un confronto tra i risultati dell'autovalutazione e i risultati del test di lettura. Nel 34% dei casi c'era una corrispondenza completa tra l'autovalutazione ed il risultato del test. Tuttavia, nel 74% dei casi l'autovalutazione divergeva dai risultati del test per non più di un livello. La corrispondenza tra i risultati del test e l'autovalutazione era più alta per i partecipanti di livello linguistico più elevato. Il numero delle persone che si sopravvalutano è più alto di quelle che si sottovalutano (il 20.6% contro il 6.2%). La sottovalutazione delle proprie competenze linguistiche è più frequente nell'autovalutazione delle partecipanti di sesso femminile rispetto a quella degli uomini.

## **Parole chiave:**

Autovalutazione, affidabilità, competenze linguistiche, livelli, Quadro Europeo di Riferimento

## **1. Context**

Assessment tools for modern foreign languages are intended to measure in a reasonably precise manner the level of linguistic proficiency (knowledge, skills and attitudes) of individuals in order to be able to predict their application in specific contexts. The tools one may wish to use are influenced by the particular contexts of language use (listening, oral production, oral interaction, reading and writing), the context in which the assessment is carried out (a school examination, on-the-job assessment, assessment of individuals or of a population as a whole) and their reliability.

In the context of the research project PNR 56 *Plurilingual competence and identity of young adults (30 years old) in the Italian speaking part of Switzerland*, we were interested in assessing linguistic competencies of former

junior secondary education students in the Swiss canton of Ticino<sup>1</sup>. Not only did we look at their current linguistic proficiency but we also analysed their development since leaving compulsory education in 1992.

For this purpose, the following types of assessment were selected:

- 1) Assessment of the linguistic proficiency of secondary school students carried out by their teachers at the end of the compulsory schooling in 1992 through traditional school methods (tests, exams, etc.).
- 2) Assessment of the role of second languages in companies, businesses and the administration (people in charge of human resources or staff development) on the level of competences of their young staff members. These results have been detailed and analysed in greater depth in the context of a focus group.
- 3) Assessment of the current linguistic proficiency of the 30 year-old Ticino population based on a representative sample of about 900 people.

For this purpose a questionnaire was developed. It contained several questions with respect to their knowledge of and competence in the Swiss national languages and in English<sup>2</sup> (actual situation and development).

## **2. Assessment of linguistic competence**

In order to assess their linguistic competences, these persons were asked to carry out a self-assessment. It would have been too time-consuming to ask the subjects to complete a full-scale test. Ideally, an instrument like Dialang ([www.dialang.org](http://www.dialang.org), Alderson & Huhta, 2005), a computer-based adaptive language testing system freely available on the internet, would be preferable but time being limited they were asked to assess their linguistic abilities on two out of the five competences of the Scale of the Common European Framework of Reference (CEFR) (Council of Europe, 2001), which is also used in the European Language Portfolio.

Participants were asked to assess their competence in reading and in oral interaction.

---

<sup>1</sup> Italian is the language commonly spoken in Ticino and this is also the language of instruction in schools; it is the mother tongue of the vast majority of the individuals in our sample, though some of them may be bilingual (in varying degrees) or have another mother tongue.

<sup>2</sup> Most individuals began learning these languages during their period of secondary schooling. English, at that time, was still an optional subject, but, with a few exceptions, all students were obliged to learn French and German. Many of them have continued their language studies either in regular school contexts, through extended stays in other parts of the country or abroad, or by taking private language courses. Over a period of fifteen years most former students have needed the languages they learnt either during their studies in other parts of Switzerland or in work-related contexts.

## 2.1 *Reliability of self-assessment*

However, there are some doubts with respect to the reliability of self-assessment. The research on the reliability of self-assessment is not conclusive. Braegger (2001) analyses the literature on self-assessment and states "Die Auswertungen dieser Studien haben gezeigt, dass die Validität der Selbstbewertung oder des Selbstkonzepts mässig mit "objektiven" Messungen übereinstimmt" (p. 56).

Other studies claim that self-assessment is quite reliable. According to Coombe (2002) "with regard to second and foreign language, research reveals an emerging pattern of consistent, overall high correlations between self-assessment results and ratings based on a variety of external criteria". Alderson (2005) reports on the aspect of self-assessment in the computer-based testing system Dialang in which self-assessment has been systematically integrated. He reports extremely high rank order correlations between CEFR level test scores and self-assessment: writing  $r=.84$ , reading  $r=.91$  and listening  $r=.87$ . These correlations apply to English and are based on 645 respondents from eleven European countries.

Alderson concludes that "there is considerable agreement between the test and self-assessment results in terms of CEFR-levels, for all skills. Of course, discrepancies may be due to under- or over-estimation on the part of the learners, or to inappropriate cut-offs for the CEFR-levels in either tests or self-assessment or both" (p. 109).

In a study conducted by Van Onna & Jansen (2006) 293 staff members of various organisations in the Netherlands were interviewed about their language competence in foreign languages. Again the Dialang system was used, but the researchers found very low correlations between self-assessment and the test scores on Dialang. The subjects in this study systematically overestimated their linguistic competence in foreign languages, whereas the subjects in Alderson's study tended to underestimate themselves. The levels of the Dutch subjects as defined as scores on the Dialang test were often at the level of beginner or lower intermediate, whereas their self-assessment was at the level of upper-intermediate or advanced.

## 2.2 *Self-assessment in language learning contexts*

There appears to be a difference in the reliability of self-assessment in language learning contexts, in which students gradually become familiar with self-assessment procedures, and situations in which self-assessment is used to collect information on the level of proficiency of a given group of people. In language learning situations, course objectives are now often stated as so-called can-do statements and when these are also used for self-assessment, learners will have had an opportunity to familiarize themselves with them. In addition, self-assessment in these cases is usually a check as to whether a

specific level has been achieved which is usually the level the course was aiming at.

### *2.3 Cultural factors influencing self-assessment*

There are also individual and cultural factors that influence self-assessment. Teachers report that uncertain students have a tendency to underestimate their skills, whereas self-confident ones tend to overestimate themselves. In addition, individuals' self-assessment may be influenced by their attitude towards the language or the people who speak it. In Van Onna & Jansen's study the Dutch subjects systematically overestimated themselves. Alderson (2005) reports significant difference across mother tongues on self-assessment reading ability. Unfortunately the sample of Italian mother tongue speakers in Alderson's sample was too small to be considered.

## **3. High-stakes and low-stakes situations**

In the context of the study by Van Onna & Jansen, the participants in the study had to install Dialang on their home computers, carry out the test and report the outcomes. This situation can be defined as a low-stakes situation (Roever, 2001). For the participants the score is not very important for either their professional or their personal lives. This may have affected the seriousness with which the test was undertaken. The higher level items in particular require a considerable degree of concentration, as anyone who has taken this test knows. If the stakes are low, participants may not have taken the test so seriously, which may have resulted in lower test scores. The situation in which the subjects in our test took the Dialang test can also be defined as low-stakes, Dialang itself being a low-stakes testing system. Subjects were either invited to take the Dialang tests at a designated centre in a computer lab (n=56) or took the test on the interviewer's laptop computer when they were visited at home (n=40). This may have had a positive effect on the seriousness with which the test items were completed. In addition, participants received a certificate that stated the level they had achieved which may have enhanced the reliability and reduced the "low-stakes effects". The impression of the team members who supervised the test was that most test-takers took the test quite seriously.

## **4. Test scores**

For our purposes we asked the test-takers in our sample to do the reading test on the Dialang system. Tests were taken for German (n=36), English (n=60) and French (n=50). Some individuals took tests for more languages, usually because they were interested in receiving a certificate. We received

146 evaluations in total, 79 by males and 67 by females. The total number of individuals was 99<sup>3</sup>.

Individuals who took the test in more than one language did not systematically overestimate or underestimate themselves, so we treated them as separate evaluations.

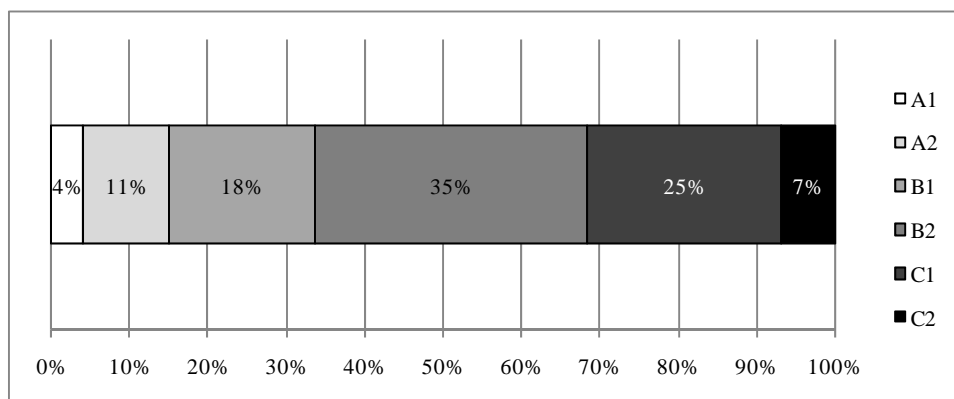


Fig. 1: Results Dialang test

More than a quarter of the individuals who agreed to take the test had a reading proficiency of an advanced level in the languages tested. Far fewer people who took the test achieved results at the A1 and A2 levels. Not many individuals who are aware of their limited linguistic competence agree to be submitted to a test voluntarily. Most individuals had language skills in the B1 and B2 field.

#### 4.1 *Relating test scores to self-assessment*

The individuals in our project assessed themselves twice, the first time as part of a written questionnaire, the second, a few months later, when taking the Dialang test. The scores on the Dialang self-assessment were registered and

<sup>3</sup> The persons who accepted to take the Dialang test belonged to the 900 person sample representative of 30 year-olds of the Ticino canton. We were therefore able to assess to a great extent the representativeness of this subsample compared to the main sample. It should be particularly noted that the Dialang subsample was characterised by a certain overrepresentation of the socio-professional middle-upper and upper categories (about 45.5% compared to the 35.4% of the main sample) as well as a subsequent underrepresentation of the other categories, particularly the lower (7.1% compared to 13%).

This situation also implies an underrepresentation of the persons who, 15 years before, had good school results (marks) compared to those who had a middle or low level. If one takes into consideration the self-evaluation (questionnaire) on the level of proficiency in the three languages by contrasting the main sample and the Dialang sample, one may note that the latter implied a certain overrepresentation of persons who self-assessed at C level of French and English (29.9% against 20.5% for English) with subsequent underrepresentation of the lower levels. For German, the overrepresentation of the B2 level is balanced by the underrepresentation of the B1 level (15.8% compared to 23.2%).

compared with the self-assessment on the questionnaires. The correlation between the two self-assessments was  $r=0.68$ ,  $p<0.01$ .

The scores on the self-assessment questionnaire were related to the scores on the Dialang test.

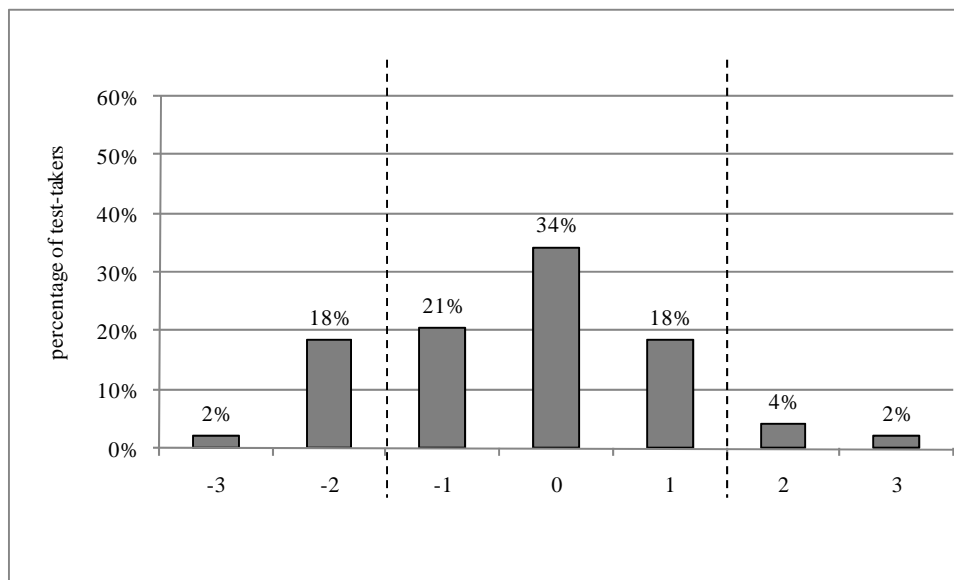


Fig. 2: Differences between Dialang test results and self-assessment

With respect to the link between the self-assessment and the test score, we see that about one third of the test-takers' self-assessment corresponds to the results of the test. 41.1% overestimated their reading ability while 24.7% underestimated their ability to read texts in a foreign language.

- Self-assessment = test score: 34.2%
- Self-assessment above test score: 41.1%
- Self-assessment below test score: 24.7%

Since the six levels of reading comprehension are fairly broad, we also looked at the results when we accepted a discrepancy of one level. The Dialang scores do not allow us to see whether a test-taker is a low or a high performer within a given level. This is due to the adaptivity of the test. With only very few additional correct or incorrect answers an individual might have scored one level higher or lower. Therefore, it is interesting to see when the discrepancy is higher by more than one (e.g. self-assessment A2, test score B2).

This analysis shows that 20.6% of the test-takers overestimated their reading competence, of which 2.1% by more than two levels (real score A2 estimated score C1). 24.7% of the individuals underestimated their reading competence, of which 6.2% by more than one level. Taken all together, we see that the self-

assessment of nearly three quarters is reasonably reliable, whereas nearly one quarter is not reliable, since the discrepancy was more than one level.

The results also show that more people tend to overestimate their proficiency than to underestimate it (20.6% versus 6.2%). This tendency was also found in the Dutch study (Van Onna *et al.*, 2006).

The highest number of correct self-evaluations occurs at the C2 level: 70%. For the other levels the figures are C1: 36.1%; B2: 39.2%; B1: 29.6%; A2: 25%. Self-assessment tends to become more reliable at the higher levels, which is not surprising because learners have more experience and are probably in a better position to assess themselves. All test-takers at the A1 level (n=6) had overestimated themselves<sup>4</sup>.

#### 4.2 Differences according to language.

There are some differences between the three languages but they are fairly marginal. The results can be seen in the following figure:

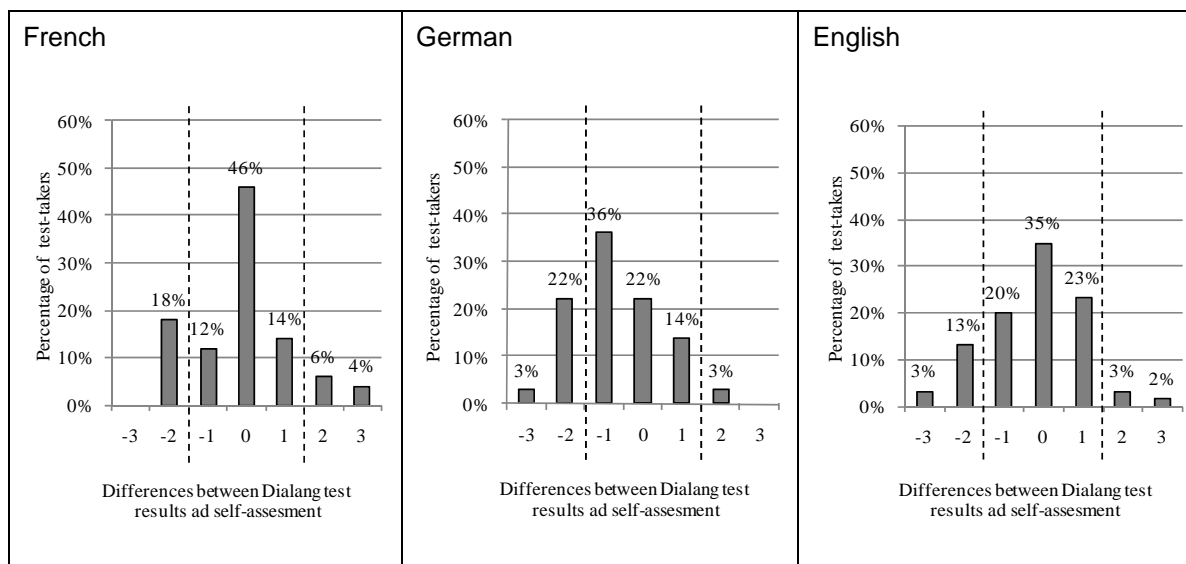


Fig.3: Differences according to language

The numbers of acceptable self-evaluations are very similar.

- French: 72.0%
- German: 72.2%
- English: 78.3%

The deviations are also irrelevant in view of the total number of test-takers.

<sup>4</sup> As a consequence of the relative overrepresentation of the good speakers in the Dialang sample compared to the main sample and even more compared to the global population under consideration, one may expect that a further slight error factor in the self-assessment can be added to the one mentioned above.

### 4.3 Differences between male and female test-takers

An analysis of results of male and female individuals shows that there is a slight tendency for females to underestimate their language competence. 10.5% of the female test-takers have underestimated their reading competence by more than one level.

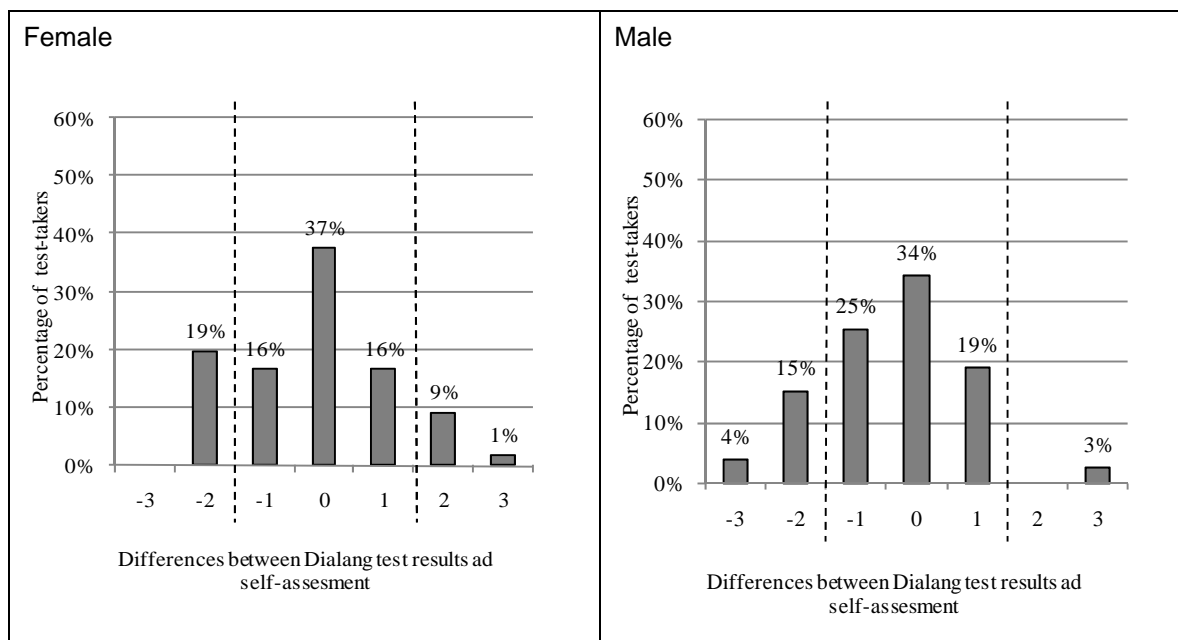


Fig. 4: Differences between male and female test-takers

## 5. Conclusion

With respect to the reliability of self-assessment on the basis of the Council of Europe's reference scale we can conclude that it is fairly reliable if one accepts a difference of one level. More than 70% of the test-takers were able to assess their reading competence in a foreign language in this way. Fully correct self-assessment was found in only one-third of the test-takers. It shows that individuals have a general awareness of their competence, but that it is not very specific with reference to the scale.

The self-assessment in our test was for reading, which is a skill for which self-assessment is easier than for speaking and writing, since it is harder for a person to assess the level of grammatical, collocational or phonological correctness. One should be aware of this issue when self-assessment for reading is used for drawing conclusions with respect to general communicative competence in all skills.

Our research confirms what is known from the scientific literature that self-assessment tends to be more reliable, when individuals have achieved a high level of linguistic competence and in educational contexts in which a culture of



self-assessment is being fostered. The context of our research is comparable with the situation in the Dutch research by Onna & Jansen (2006), in which low correlations between self-assessment and Dialang scores were found.

The conclusion that detailed self-assessment scores in non-educational contexts should be interpreted with care seems justified.

## REFERENCES

- Alderson, J. C. (2005): Diagnosing foreign language proficiency. The interface between learning and assessment. London (Continuum).
- Alderson, J. C. & Huhta, A. (2005): The development of a suite of a computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22 (3), 301-320.
- Braegger, M. (2001): Wie schätzen Schüler/Innen ihre (Fremd-)sprachenkenntnisse ein. Preisgekrönter. Arbeit im Rahmen eines Wettbewerbs im Europäischen "Jahr der Sprachen" der Universität Basel.
- Coombe, C. (2002): Self-assessment in Language Testing: Reliability and Validity Issues. Available: <http://www.philselfsupport.com/selfassessment.htm> (20.12.2008).
- Onna, B. van & Jansen, C. (2006): Nederland talenland? Over de beheersing van Engels, Duits, Frans en Nederlands in Nederlandse organisaties. (The Netherlands, a country of foreign languages? On the mastery of English, German, French and Dutch in Dutch organizations). In: B. Hendriks, H. Hoeken & P. J. Schellens (eds.), *Studies in Taalbeheersing*. Assen (Van Gorcum).
- Roever, C. (2001): Web Web-based language testing. *Language Learning & Technology*, 5(2), 84-94. Available: <http://llt.msu.edu/vol5num2/roever/default.html> (22.11.2008).
- Saito, Y. (2003): The use of Self-assessment in Second Language Assessment. Available: [http://www.tc.columbia.edu/academic/tesol/WJFiles/pdf/Saito\\_Forum.pdf](http://www.tc.columbia.edu/academic/tesol/WJFiles/pdf/Saito_Forum.pdf) (22.12.2008).