Self-assessment vs. external evaluation of proficiency in English as a foreign language among student teachers

Carlos L. Salvadores Merino. Universidad de Granada. Grupo de investigación HUM-840 "Conocimiento Abierto para la Acción Social"

Correspondencia a través de **ORCID**: Carlos L. Salvadores   iD   **0000-0002-6991-8309**

Abstract: One of the aims of Higher Education is the promotion of Lifelong Learning and Self-Assessment (SA) could be considered as one of its main pillars. However, there are serious doubts about its reliability and validity as an evaluation tool. The aim of this correlational study is to explore SA accuracy contrasted to objective tests by comparing the results of the correlations between the SA tests conducted by 46 university (L2) English learners and their external evaluation, marked by teaching staff, for the purpose of testing the accuracy of_SA. The SA tool used in the study was the Biography element of the European Language Portfolio (Form 3 of the ELP Passport for Adults); and that of the external evaluation was a model based on the Cambridge FCE tests administered by the UCLES. The results revealed that correlation on the overall scores were high, demonstrating that the students' own assessments of their proficiency in the English language were accurate, being higher in the oral skills but lower in the written skills. The findings also revealed higher correlations in the expression skills than in the comprehension skills. This study contributes to affirm the accuracy and validity of SA as an evaluation instrument.

Keywords: self-evaluation, English as a foreign language.

Autoevaluación frente a la evaluación externa de la competencia en lengua inglesa como lengua extranjera en estudiantes de magisterio

Resumen: Uno de los objetivos de la Educación Superior es la promoción del aprendizaje a lo largo de la vida y la autoevaluación puede considerarse como uno de sus pilares fundamentales. Sin embargo, existen serias dudas sobre su fiabilidad y validez como herramienta de evaluación. El objetivo de este estudio correlacional es explorar la precisión de la autoevaluación contrastada con las pruebas objetivas comparando los resultados de las correlaciones entre las pruebas de autoevaluación realizadas por 46 estudiantes universitarios de inglés como lengua extranjera y su evaluación externa, corregida por el profesorado, a fin de comprobar la precisión de la autoevaluación. Como instrumento de autoevaluación se utilizaron los impresos de la Biografía del PEL y para la evaluación externa un modelo tipo de las pruebas del FCE administradas por UCLES. Los resultados globales mostraron altos coeficientes de correlación, demostrando que los estudiantes fueron precisos en la autoevaluación de su dominio de la lengua inglesa, siendo éstos mayores en las destrezas orales pero más bajos en las escritas. Los resultados también revelaron correlaciones más altas en las destrezas de expresión que en las de comprensión. Este estudio contribuye a confirmar la precisión y validez de la autoevaluación como instrumento de evaluación.

Palabras clave: autoevaluación, inglés como lengua extranjera**.

Introduction

Over the last few decades (Peirce, Swain & Hart, 1993), there has been growing recognition of the value of self-assessment in teaching and learning processes (Bachman & Palmer, 1989; Blanche & Merino, 1989 Davidson & Henning, 1985; Blue, 1988; Henner-Stanchina & Holec, 1985; Janssen-van Dieten, 1989; LeBlanc & Painchaud, 1985; Oscarsson, 1989, 2013). Interest in this topic is largely due to the development of student-centred language teaching methodologies and autonomous learning (Brown & Hudson, 1998; Dickinson, 1987; Holec, 1980; Knowles, 1975; Nunan,1988; Rea, 1981; Riley, 1985) and to the development of professional

competences (Finnie & Meng, 2005; Sundstroem, 2004). Thanks to the publication of the Common European Framework of Reference for Languages (CEFRL) (CoE, 2001,2018) and the European Language Portfolio (ELP), self-assessment now has a pivotal role in foreign language teaching and learning (Little,2005). It promotes personal reflection and self-criticism during the learning process, enabling students to analyse both the process and the product, establish what they know, and, above all, identify how they might expand that knowledge (García *et al.*, 2004).

A further reason for the prominence of self-assessment is the critical role it has played in lifelong learning, one of the key objectives of the European Learning Portfolio (Dobson,2006:27). Boud and Falchicov (2006) and Dochy, Segers and Sluijsmans (1999) concur that Higher Education should contribute to the promotion of both lifelong learning and self-assessment, as traditional assessment methods are not suited to the aims of this kind of learning: reflexive thinking, critical attitude, capacity for self-assessment and problem-solving capacity. Equally, Boud *et al.* (2013) note that developing the capacity among students to conduct self-evaluations of behaviour is widely assumed to be one of the ultimate objectives of Higher Education.

Theoretical framework

This interest in self-assessment has led to growing research on the topic. Studies on self-assessment have tended to fall into one of two types, according to their aims (Oscarsson, 1984, cited in Finch, 2003: 6): on the one hand, research on the different ways in which the student can participate in their own evaluation (Finch, 2003; Harris, 1997;Heilenman, 1990; LeBlanc & Painchaud, 1985;Oscarsson, 1978, 1984; Von Elek, 1982;); and on the other hand, research on the extent to which the different self-assessment methods and tools are able to deliver relevant and reliable results (Alderson, 2005;Baleghizadeh & Hajizadeh, 2014; Brantmeier, Vanderplank, & Strube, 2012; Bachman & Palmer, 1989; Cheng, 2008; Dochy et al., 1999;Edele *et al.*, 2015; Finnie & Meng, 2005; Ma, W., & Winke, P. 2019; Runnels, 2016). These works (in the main, correlative studies), establish comparisons between the results of self-assessment and those deriving from different types of external tests, such as standardized tests or proficiency tests. These studies endeavour to establish whether the results of self-assessment processes are accurate and whether their level of validity and reliability is therefore acceptable. However, they have arrived at different and contradictory conclusions on this question. On the one hand, we find studies whose results from students' self-assessments have generated a degree of scepticism and lack of confidence in their ability to evaluate their own achievements, particularly in formal educational contexts (Blue,1988:100;Davidson & Henning, 1985; Janssen-van Dieten,1989:31; Peirce et al.,1993:38).The most common causes of the low correlations were found to be that the students had received no training or guidance in self-assessment (Harris, 1997:19), or that there was a mismatch between the content of the self-assessment instrument and that of external tests (Runnels, 2016). Other studies find only moderate correlations (Alderson, 2005; Brantmeier et al., 2012; Edele et al., 2015). Their authors consider the accuracy of skills in self-assessment to be questionable, mainly in view of the quality and disparateness of the tools.

On the other hand, there are several studies that claim the results of the self-assessments proved them to be highly reliable (Bailey, 1998; Baleghizadeh & Hajizadeh, 2014; Blanche & Merino, 1989;Boud & Falchikov, 1989;Delgado, Guerrero, Goggin & Ellis, 1999; Dochy et al., 1999;LeBlanc & Painchaud, 1985; Oscarsson, 1978, 1989; Ross, 1998: 10; Wilson, 1999;). These authors concur that the statistical results of their studies on self-assessment demonstrate the validity and reliability of this method, presenting correlation coefficients with values typically in a range of r = .50 to

.60 (Blanche & Merino, 1989: 315), r = .53 (LeBlanc & Painchaud,1985:679); r = .63 (Ross, 1998: 4) or r = .58 to .64 (Bailey, 1998, cited in Brown, 2004: 270).

Research aim

The aim of the present exploratory study was to analyse the degree of accuracy of self-assessment results in foreign language proficiency among student teachers. The research questions were as follows:
Are there correlations between external evaluation and self-assessment, in overall terms?
Do these correlations also exist across the different, specific language skills?
What might such correlations at the individual skill level tell us about the factors that may be influencing the relationships?

Methodology

Population and sample

The study sample comprised 46 student teachers (n = 46), mostly aged 19 to 22, from the Faculty of Education and Humanities Campus of Ceuta (University of Granada). Of these, eleven were studying the first year of the Foreign Language (English) specialization. The remaining 35 were second-year students in other specializations (Early Childhood Education, Physical Education, Music Education and Primary Education), who were on the Foreign Language Teaching (English) course at the Faculty. The sample overall comprised 13 men and 33 women. As there are a relatively small number of students at this faculty, no pre-selection was undertaken. All students enrolled at the time, within the target population, were simply invited to take part in the study. In terms of response, out of a total of 93 such students, 46 opted to take part – approximately 50% of the population in question.

Despite the relatively small sample, we believe the results offer valuable insights, given that correlational methods are considered a powerful exploratory tool that do not rely on large sample size (Cohen & Manion, 1990).

Evaluation tools

For the self-assessment tool we chose the table of descriptors for self-assessment taken from the ELP Biography for Adults (tick-box Forms 8.1–8.6) (Council of Europe). The table is divided into five skills that match those set out within the CEFRL (MECD, 2002). The participants were also asked to score themselves on the overall scale of CEFRL levels (Form 3 of the ELP Passport for Adults), by ticking the box corresponding to the level they believed they had achieved (across all five linguistic skill-sets and also in an overall score for the self-assessment test).

For the external evaluation instrument, we selected the standardized test for the First Certificate in English (FCE) administered by the University of Cambridge Local Examinations Syndicate (UCLES). This test comprises the following elements: Reading Comprehension, Writing, Use of English, Listening Comprehension and Speaking. As in the self-assessment, an overall score for each student's FCE test result was noted.

All the tests were conducted during a two-week period approximately, the self-assessments being undertaken first, followed by the FCE tests. The ELP self-assessment questionnaires were completed under the supervision of the author of the present research. As this was the first time the participants had ever undertaken a test

of this kind, this supervision ensured they all completed the process correctly. The FCE tests were conducted in line with the formal instructions presented with the different exams, regarding questions of format, marking criteria, qualifications criteria and timing. The tests were then marked by the author of this research under the supervision of two accredited external UCLES examiners.

Correlational study

This is an exploratory study whose purpose was to examine whether there were correlations between the scores obtained by students in their self-assessment tests, measuring their own proficiency in English as a foreign language, and those obtained through the use of external evaluation tests, marked by teaching staff. To this end, Pearson's Product-Moment Correlation was used. The subjects were all studying for their teacher training qualifications in Higher Education (in a similar vein to the aforementioned studies, which focused on university students who were learning a foreign language) (Edele et al., 2015: 101).

We selected the type of study on the basis of the literature dealing with foreign language proficiency assessment, in which correlational studies are the most prolific (Boud & Falchikov, 1989; Oscarsson, 1989; Blanche & Merino, 1989; Dochy et al., 1999; Liu, H., & Brantmeier, C. 2019). From the studies by Ross (1998), Sundstroem (2005) and Edele et al. (2015) and their analysis of different spheres in which self-assessment is used, we also derived that in the comparison of the results obtained via self-assessment and fluency tests, correlational studies were by far the most popular choice in the literature.

Resultados

Borra y escribe los resultados de forma concisa y clara. Se redactarán de modo que no haya duplicación ni repetición innecesaria de información en el texto, figuras o tablas.

Research results

The students in our sample presented a great variety of levels in terms of English language proficiency (as it comprised both specialist students taking the Foreign Language–English specialist track and also non-specialists). Nevertheless, the results themselves presented a high degree of homogeneity.

Overall results of the self-assessment test (ELP questionnaire)

To conduct our statistical analysis, we assigned values to the different levels of the CEFRL scale: a value of 1 to level A1, a value of 2 to level A2, and so on, up to a value of 6 (C2). We should point out that this communicative competence scale is not based on equidistant intervals; rather, the continuum is divided into unequal tranches (hence, level A1 covers less distance on the proficiency continuum than level A2). In other words, the more proficient the individual becomes, the higher up the scale they progress and the greater the intervals between levels (Savignon,1983). However, for the purposes of our investigation, we assigned all the levels the same proportion of the continuum, thus enabling the relevant statistical analyses to be conducted.

Table 1. Values assigned for the purposes of statistical analysis in the self-assessment test.

Value	CERFL Level

1	A1
2	A2
3	B1
4	B2
5	C1
6	C2

This analysis provided the following results: first, the overall average score from the students' self-assessment was 2.43 (out of a possible maximum of 6), indicating that they had achieved, on average, level A2 on the CEFRL scale. This is the level that, according to Spanish national regulations (art 3.6,R. D. 1629/2006, de 29 de diciembre), Baccalaureate graduates (similar to the UK A-Level) should have achieved. Second, the range of levels achieved in the self-assessment test varied between A1 and B2, while none of the participants assessed themselves as having achieved the level of "Proficiency" or "Mastery" (C1 or C2). Two-thirds of the sample (69.5%) averaged scores corresponding to levels A2 (47.8 %) and B1 (21.7 %). This result indicates that the sample, with a standard deviation of 0.93 and a mean and mode of 2, respectively, presented a homogeneous distribution and that there were very few extreme cases that might weaken or skew the results of the correlations (Bachman, 2000: 98). The lowest and highest levels – A1, accounting for 13% and B2, 17.4% – represent just one third of the entire sample. As these percentages are similar, the distribution curve is also homogeneous at the two extremes.

 Overall results of the externally-evaluated test (FCE test)

The average score from across the sample in the external evaluation (based on the FCE/UCLES standardized tests) was 37.76, which represents 62.93% of the pass-rate score of 60. The median was 34.95, the mode was 27.40 and the standard deviation was 17.70. With this test, it is not possible to establish such a direct comparison as we were able to with the results of the self-assessment. In the absence of clear, explicit points of reference, the possible correspondences in the results of the FCE test can only be inferred. On the premise that the pass rate is 60, if we imagine a continuum of linguistic proficiency, then level A2 in the CEFRL framework (which is the mid-point between A1 and B2) equates to approximately 30 points in the FCE test. We can therefore infer that the average score achieved by the sample in the external evaluation (37.76) equates to the level of proficiency designated A2.

Table 3. Summary of overall linguistic proficiency level (FCE) (n=46)

	Listening	Speaking	Writing	Reading	Use of English	Overall FCE level
Mean	8.92	9.04	5.65	9.95	4.21	37.77
Median	8.35	8	4.30	9.80	3.48	34.95
Mode	9.30	8	.60	12.30	2.40	27.40
Standard dev.	3.95	4.29	4.86	3.97	2.98	17.70

Overall results of the language skills ELP

Starting with the self-assessment (ELP) results according to specific linguistic skills, here, with only slight variations, a level of A2 was achieved, which is the same level as that registered for the *overall* proficiency results (based on the responses to ELP Form 3). This indicates that the students perceived their level of proficiency similarly, whether in the general sense or across the different linguistic skill-sets. Specifically in Reading and Listening Comprehension, the median and the mode were slightly higher; while skills relating to self-expression (Speaking and Writing) were rated slightly lower.

Table 2. Summary of overall linguistic proficiency level (ELP) (n=46)

	Listening	Speaking	Interactive comm.	Writing	Reading	Overall ELP level
Mean	2.36	2.50	2.45	2.36	2.91	2.43
Median	3.00	2.00	2.00	2.00	3.00	2.00
Mode	3.00	2.00	2.00	2.00	3.00	2.00
Standard dev.	1.08	.96	1.02	.951	.93	.93

This self-perception of greater proficiency in the Comprehension skill-set was affirmed in the results of the FCE test, where the highest test scores were achieved in Reading and Listening comprehension skills, respectively. Again, it was in the self-expression skill-set that the lowest scores were obtained (for Speaking and Writing, respectively), albeit with greater differences.

In summary, then, in both the ELP and the FCE tests (self-assessment and external evaluation), the higher scores were obtained for comprehension skills, compared to self-expression, and for written skills, compared to oral skills.

One particularly interesting feature was that of the low scores awarded for Use of English, which were markedly poorer than scores for all the other skills. Of the 46 students in the sample, only three achieved a score of 10 or above, out of a possible 20 – a striking difference. However, an earlier study by Jiménez Jiménez (2004) on the level of English proficiency among Spanish student teachers also showed that results for Use of English were noticeably lower than those for the other linguistic skills.

Correlations in the overall scores obtained in the ELP and the FCE

The first question to be addressed is whether a correlation can be established between the two sets of data without having to apply any previous treatments. The FCE scale ranges from 0 to 100 and the skill tests range from 0 to 20 (based on the respective transformations we made, as described earlier), while the ELP ranges from 0 to 6. We deemed it unnecessary to make any transformations, as both variables operate in the same direction (the higher the ELP value, the higher the score; and the higher the FCE value, the higher the score). In other words, the scales themselves are not important – what matters is the direction of the variables, which enables them to be interpreted. Correlation measures a) the degree of association that exists between two variables, b) their trend and c) whether it is possible to predict the behaviour of one via the other. Thanks to its mathematical definition, the scale on which the correlation is measured is

of no importance. Brantmeier et al. (2012) also used correspondences, as the scales for self-assessment and external evaluation were different.

The second question is: if correlations are indeed established, how might the coefficients be best defined? Sometimes, absolute values (varying between values of 0 and 1) are described as weak, moderate or strong, depending on how close or far they are from 1. Such definitions are arbitrary as they rely, in part, on the variables under study, the sample size and even the expectations of the researcher (Gardner, 2003: 180). In the present study, for our reference point we used the definition established by Cohen (1977, cited in Falchicov & Boud, 1989: 420), which provides an operational definition of values for *r*. On this basis, a value of *r* = 0.10 is *low*; a value of *r* = 0.30 is *moderate;* and a value of *r* = 0.50 is *high.* Bachman (2004: 103) similarly considers that a coefficient of *r* = .44 is relatively high.

We can observe that the results of both tests produce correlations with a high Pearson Coefficient of *r* = .651, and that these correlations are significant (*p* = .000). We may therefore affirm that there are indeed correlations between external evaluation and self-assessment. However, we must acknowledge that, due to the relatively small sample size, these results should be interpreted with caution and in light of their context. That said, we believe they are indicative of the existence of correlations, albeit perhaps slightly less strong.

Turning now to the correlations between the scores for linguistic skills (ELP and FCE), the following results were obtained:
Listening: this presents a correlation with a high and significant coefficient (*r* = .581 and *p* = .000, respectively)
Writing: here too we find a high and significant coefficient (*r* = .510 and *p* = .000)
Reading: this also presents a correlation with a medium–high and significant coefficient (*r* = .397 and *p* = .006)
Speaking: here we also find a correlation with a high and significant coefficient (*r* = .690 and *p* = .000).

With regard to linguistic skills, then, we can see that these also produce correlations with high coefficients. The highest were found in Speaking (*r* =.690), followed by Listening (*r* = .581). At the other end of the score range, we find the skills related to expression and written comprehension (that is, Writing and Reading), with coefficients of *r* = .510 and *r* = .397, respectively. These results largely coincide with the findings of Blanche and Merino (1989), and of Ross (1998). They found correlations with high coefficients not only between both types of evaluation, but also between linguistic skills (albeit with small variations).

Table 4 Correlation coefficients for linguistic skills, compared to those of Ross (1998)** p = .01

Linguistic skill	Present research	Ross (1998)
Listening	*r* = .581**	*r* = .650
Writing	*r* = .510**	*r* = .524
Reading	*r* = .397**	*r* = .613
Speaking	*r* = .690**	*r* = .554

In Ross's study (1998), the correlation coefficients present few differences, the highest corresponding to comprehension skills (Listening and Reading). In the present research, the differences are slightly more prominent. The highest coefficients appear in the oral skills, Speaking and Listening (r = .690 and r = .581, respectively). By contrast, the written skills, Writing (r = .510) and Reading (r = .397) present lower coefficients. If we compare the comprehension skills (Listening and Reading) with self-expression skills (Speaking and Writing), we observe that the coefficients are higher for the latter skill-set. Two conclusions can be drawn from all of these data. First, in contrast to the findings of Ross (1998), the students in our sample obtained higher correlation coefficients in their self-expression (either oral or written) than in their comprehension skills. And second, these coefficients are higher for oral than for written skills. Similarly, in the study undertaken by Runnels (2016), the correlation indices for Listening were moderate, while those for Reading were weak.

Discussion

One of the difficulties related to data analysis in correlational studies is the heterogeneous results found in many of these studies, ranging from very low to very high correlation coefficients, due to a large variety and quality of the instruments used, sample sizes or linguistic dimensions assessed (Edele *et al.*, 2015), which should always be taken into consideration before data analysis.

In this study, significant correlation with a high coefficient (r = .651 and p = .000) was found between the overall scores from the ELP questionnaire and the overall results of the FCE test, which is in line with the findings of other self-assessment correlational studies (Bachman & Palmer, 1989; Blanche & Merino, 1989; Delgado et al., 1999; Fimmie & Meng, 2005; LeBlanc & Painchaud, 1985; Oscarsson, 1989; Ross, 1998; Stefani, 1994; Wilson, 1999;). In these studies, the correlation coefficients range between r = .50 and .60, thus answering the primary question posed by the present research.

We can therefore affirm that the self-assessments produced by students participating in our study, regarding their proficiency in the English language, were accurate. One reason for this high correlation may be that the self-assessment tool used in the study – the ELP questionnaire – is based on everyday situations, contains a large number of items and uses specific criteria presented in great detail (Edele et al., 2015). Brantmeier et al. (2012) and Ross (1998) suggest that a high level of specificity improves the accuracy of self-assessment, as items expressed as "can do" (such as those in the ELP) are more accurate than other possible formats. Equally, when the questions in the self-assessment are specific rather than generalized, and the scale is absolute, the results tend to be more realistic (Ackerman et al., 2002, cited in Sundstroem, 2005).

The sample composition may provide a further reason for the high correlation in the present study. These were students of a foreign language, as opposed to the populations of immigrants used in other studies, such as those of Finnie & Meng (2005) and Edele *et al.* (2015). Students tend to be more accurate in their self-assessments, thanks to the feedback they receive on their linguistic skills

during their studies, and their frame of reference is less ambiguous (Edele *et al.*, 2015). They also possess broad experience of learning, together with knowledge of the skills being evaluated (Ross, 1998), which has a positive impact on the accuracy of their assessment.

Our second research question was: do correlations also arise across the different specific language skills? Once again, the answer is "yes". Significant correlations with high coefficients were found in almost all of the linguistic skill-sets.

The highest coefficient was found in Speaking ($r =.690$), followed by Listening ($r = .581$), Writing ($r = .510$) and, lastly, Reading ($r = .397$). The highest coefficients derive from oral skills, expression and comprehension, in that order, while the lowest pertain to the written skills of expression and, lastly, comprehension.

In the oral vs. written language dyad, the students participating in this study presented greater accuracy in their self-assessment of their oral skills than their written skills. And in the other dyad – comprehension vs. expression – they present greater accuracy in their self-assessment of the latter skill-set. These results contradict those of Peirce et al. (1993: 35), who found that Reading attracted a higher correlation coefficient, as students had greater confidence in their comprehension-related skills than in their self-expression skills, because the former are more developed than the latter. Ross (1998) also found that listening and reading comprehension skills produced the highest correlation coefficients. He attributed this result to the broad experience among foreign language students in this area, particularly at university level (as in the present study). Oscarsson (2013) considers that prior experience is also a key factor as well. Edele et al. (2015) assert that Listening and Reading skills are strong indicators of linguistic proficiency overall, as they are essential for following classroom content. Hence, their mastery of comprehension, coupled with the greater confidence felt by these students, would explain the stronger self-assessment results in comprehension skills identified by these authors.

However, in the present study, the correlation coefficients for the self-expression skills are higher than those of the comprehension skills. Specifically, we find that the oral expression skill, Speaking, presents the correlations with the highest coefficients, while written comprehension (Reading) presents the lowest coefficients. On the premise that much of the self-assessment literature (Boud & Falchikov, 1989; Falchikov & Boud, 1989; Oscarsson, 2013; Ross, 1998; Mae & Winke, 2019) holds that, the greater the student's mastery and experience, the more accurate their self-assessment results, the results of the present investigation are somewhat contradictory. Here, the students have more learning experience in written comprehension (Reading), yet they present the lowest correlation coefficients for this skill-set. The highest correlation coefficients are found in one activity in particular, Speaking, in which these students are less experienced. This activity demands greater effort, both in using this skill and in assessing the level of achievement, as it can only be assessed *post hoc.*

One possible explanation for this result of our study may be found in the very nature of oral skills. Blanche and Merino (1989) conclude that students are more accurate in their self-assessment of purely communicative skills. Elsewhere, Bachman and Palmer (1989) find that students are more aware of their skill in those areas they consider more difficult. Self-expression skills require a greater degree of foreign language mastery, as they require the student to take a more active role and carry out more complex language tasks that demand pre-planning and the implementation of production strategies (Ross,1998:7). Consequently, we might deduce that, if self-expression skills are deemed by the students to be more difficult, their perception of their level of mastery in those skills may also be more accurate.

Finally, as stated above, this is an exploratory study with its own limitations. First, the size of the sample is reduced (*n*=46), although some correlational studies have also small samples (Brown *et al.,* 2014; Dolosic *et al.,* 2016; Chen, 2008; Malabonga *et al.*, 2005). Second, albeit the ELP is considered a valid and a reliable instrument (Little, 2002; Mirici, 2008; Román & Soriano, 2015).), yet, some of the instruments contained in it, the self-assessment tool used, Form 3 of the ELP Passport for Adults, might have limited validity as it has not been researched enough, however, an inventory with specific questions about functional skills, as in the self-assessment grid in Form 3, enhance accuracy (Ross,1998). For this reason, these limitations should be considered when interpreting findings.

Conclusion

The present work found significant correlations with high coefficients between the scores achieved in self-assessment and those awarded via external evaluation (in both overall results and by specific linguistic skill). In identifying such accuracy in students' self-assessment, the work could make an important contribution to the literature by affirming the findings of previous studies that point to the accuracy and value of self-assessment.

Furthermore, the use of self-assessment as a tool for evaluation is only meaningful if the results it delivers are reliable (Edele et al., 2015). In future studies, it would be interesting to examine the factors that may influence the degree to which correlation is achieved between self- and external evaluation: proficiency level of the learners (Brantmaier *et al.*, 2012) prior experience, task specification (Edele, 2015; Ma & Winke,2019), lack of training , cultural background(Falchikov & Boud, 1989; Ross,1998), self-assessment formats and item description (Brown *el al*. 2014) so as to further heighten the accuracy of the former method.

References

Alderson, C.J. & Huhta, A. (2005) The Development of a Suite of Computer-Based Diagnostic Tests Based on the Common European Framework. *Language Testing,* 22, 301-320.

Ackerman, P.L., Beier, M.E. & Bowen, K.R. (2002) What we really know about our abilities and our knowledge. *Personality and Individual Differences,* 33, 587-605.

Bachman, L.F. & Palmer, A.S. (1989) The construct validation of self-ratings of communicative language ability. *Language Testing,* 6 (1), 14-29.

Bachman, L.F. (2000) Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing,* 17, 1-42.

Bachman, L.F. (2004) *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.

Bailey, K.M. (1996) Working for washback: a review of the washback concept in language testing. *Language Testing,* 13, 257-279.

Baleghizadeh, S. & Hajizadeh, T. (2014) Self- and Teacher-Assessment in an EFL Writing Class 1, *Gist Education and Learning Research Journal,* 8(8), 99-117.

Blanche, P. & Merino, B. (1989) Self-Assessment of Foreign-Language Skills: Implications for Teachers and Researchers. *Language Learning: A Journal of Applied Linguistics*, 39, 313-340.

Blue, G.M. (1988) Self-assessment – the limits of learner independence. In A. Brookes (ed.) *Individualisation and Autonomy in Language Learning.* ELT Documents 131. London: Modern English Publications/British Council.

Boud, D. & Falchicov, N. (1989) Quantitative studies of students' self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18, 529-549.

Boud, D. & Falchicov, N. (2006) Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31, 399-413.

Boud, D., Lawson R. & Thompson, D. (2013) Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education* , 38:8, 941-956, DOI: 10.1080/02602938.2013.769198

Brantmeier, C., Vanderplank, R., & Strube, M. (2012) What about me? Individual self-assessment by skill and level of language instruction. *System*, 40(1), 144-160.

Brown, J.D. & Hudson, T (1998) The Alternatives in Language Assessment. *TESOL Quarterly*, 32, 653-675.

Brown, D.H. (2004) *Language Assessment: Principles and Classroom Practice*. New York: Pearson Education.

Brown, N. A., Dewey, D. P., & Cox, T. L. (2014). Assessing the validity of can-do statements in retrospective (then-now) self-assessment. *Foreign Language Annals*, *47*(2), 261-285.

Chen, Y. (2008) Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12, 235-262.

Cohen, J. (1977) *Statistical Power Analysis for the Behavioral Sciences* (revised edition). New York: Academic Press.

Cohen, & Manion, (1990) *Métodos de Investigación Educativa*. Madrid: Ediciones la Muralla.

Davidson, F. & Henning, G. (1985) A self-rating scale of English difficulty: Rasch scalar analysis. *Language Testing,* 2; 164-169.

Council of Europe. (2001) *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. (2018). *Common European Framework of Reference for Languages: learning, teaching, assessment companion volume with new descriptors*. Strasbourg: Council of Europe. Retrieved from https://rm.coe.int/cefr-companion-volume-with-new-descriptors2018/1680787989

Delgado, P., Guerrero, G., Goggin, J. & Ellis, B. (1999) Self-Assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioural Sciences*, 21 (1), 34-46.

Dickinson, L. (1987) *Self-Instruction in Language Learning*. Cambridge: Cambridge University Press.

Dobson, A. (2006) Algunos aspectos de las políticas del Consejo de Europa para la educación lingüística: el Marco Común Europeo de Referencia y el Portfolio Europeo de las Lenguas. In *El Portfolio Europeo de las Lenguas y sus Aplicaciones en el Aula*. Madrid: Subdirección General de Información y Publicaciones del MEC.

Dochy, F., Segers, M. & Sluijsmans, D. (1999) The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education,* 24, 331-350.

Dolosic, H. N., Brantmeier, C., Strube, M., & Hogrebe, M. C. (2016). Living language: Self-assessment, oral production, and domestic immersion. *Foreign Language Annals*, *49*(2), 302-316. Edele, A., Seuring, J., Kristen, C. & Stanat, P. (2015) Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, 52, 99-123.

Falchikov, N. & Boud, D. (1989) Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research,* 59, 395-430.

Finch, A. (2003) Reflective Instruments for Self-Assessment in Korean EFL Classrooms. *The Korean TESOL Journal,* 6, 63-86.

Finnie, R. & Meng, R. (2005) Literacy and labour market outcomes: Self-assessment versus test score measures. *Applied Economics*, 37(17), 1935-1951.

García, F., Guillén, C., González Piñero, M., González Porto, J., Serna, I. & Vez, J.M. (2004) Marco y Portafolio: Porta Linguarum para los Europeos, *Porta Linguarum*, 2, 69-92.

Gardner, R.C. (2003) *Estadística para Psicología Usando SPSS para Windows.* Traducción de Roberto Escalona García. Pearson Educación: México.

Harris, M. (1997) Self-assessment of language learning in formal settings. In *ELT Journal* Vol. 51/1, pp. 12-20. Oxford: Oxford University Press.

Heilenman, L.K. (1990) Self-assessment of second language ability: the role of response effects. *Language Testing*, 7, 174-201.

Henner-Stanchina, C. & Holec, H. (1985) Evaluation in an autonomous learning scheme. In P. Riley (ed.) *Discourse and Learning*. London: Longman.

Holec, H. (1980) *Autonomy in Foreign Language Learning*. Oxford: Pergamon.

Jansen-van Dieten, A. (1989) The development of a test of Dutch as a foreign language: the validity of self-assessment by inexperienced subjects. *Language Testing,* 6, 30-46.

Jiménez Jiménez, M.A. (2003) *La competencia lingüística de los alumnos de la titulación de Maestro- Especialidad de Lengua extranjera (Inglés). Estudio de casos*. Unpublished doctoral thesis, University of Granada.

Knowles, M.S. (1975) *Self-directed Learning.* New York: Association Press.

LeBlanc, R. & Painchaud, G. (1985) Self-Assessment as a Second Language Placement Instrument. *TESOL Quarterly*, 19 (4), 673-686.

Liu, H., & Brantmeier, C. (2019). "I know English": Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. *System*, *80*, 60-72. https://doi.org/10.1016/j.system.2018.10.013

Little, D. (2002). The European Language Portfolio: Structure, origins, implementation and challenges. *Language Teaching, 35*(3), 182-189. doi:10.1017/S0261444802001805

Little, D. (2005). The Common European Framework and the European Language Portfolio: involving learners and their judgements in the assessment process. *Language Testing*, *22*(3), 321–336. https://doi.org/10.1191/0265532205lt311oa

Ma, W., & Winke, P. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time?. *Foreign Language Annals*, *52*(1), 66-86. https://doi.org/10.1111/flan.12379

Malabonga, V., Kenyon, D. M., & Carpenter, H. (2005). Self-assessment, preparation and response time on a computerized oral proficiency test. *Language Testing*, *22*(1), 59-92.

Ministerio de Educación, Ciencias y Deporte (2002) *Marco Común Europeo de Referencia de la Lenguas: Aprendizaje, Enseñanza y Evaluación*. Madrid: Secretaría General Técnica del MECD y Grupo Anaya.

Mirici, I. H. (2008). Development and validation process of a European language portfolio model for young learners. *Online Submission*, *9*(2), 26-34Nunan, D. (1988) *Syllabus Design*. Oxford. Oxford University Press.

Oscarsson, M. (1978) *Approaches to Self-assessment in Foreign Language Learning*. Council of Europe: Pergamon Press.

Oscarsson, M. (1984) *Self-Assessment of Foreign Language Skills: a Survey of Research and Development Work*. Strasbourg: Council of Europe.

Oscarsson, M. (1989) Self-assessment of language proficiency: rationale and implications. In RELC Journal, 19 (1), 75-93.

Oscarsson, M. (2013). Self-Assessment in the Classroom. In The Companion to *Language Assessment, A. J. Kunnan (Ed.).* doi:*10.1002/9781118411360.wbcla046*

Peirce, B.N., Swain, M. & Hart, D. (1993) Self-assessment, French immersion and locus of control. *Applied Linguistics*. 14 , 25-42.

Portfolio Adultos (de 16 o más años) - Modelo acreditado 59.2004. PEL Adultos impresos 2004, available from http://sepie.es/doc/portfolio/pelAdultosimpresos2004.pdf, accessed 24 september, 2017

Real Decreto 1629/2006, de 29 de diciembre, por el que se fijan los aspectos básicos del currículo de las enseñanzas de idiomas de régimen especial reguladas por la Ley Orgánica 2/2006, de 3 de mayo, de Educación, num. 4,pp 465-473, available from https://www.boe.es/boe/dias/2007/01/04/pdfs/A00465-00473.pdf, accessed 20 October, 2017

Román, S. N., & Soriano, S. A. (2015). Autonomous Learning and Self-Assessment through the European Language Portfolio (elp): A Pilot Study on Primary Education. *The Journal of Language Learning and Teaching*, *5*(2), 37-53.

Ross, S. (1998) Self-assessment in second language testing: a meta-analysis and analysis of experimental factors. *Language Testing*, 15 (1) 1-20.

Runnels, J. (2016) Self-Assessment Accuracy: Correlations between Japanese English Learners' Self-Assessment on the CEFR--Japan's Can Do Statements and Scores on the TOEIC®. *Taiwan Journal of TESOL*, v13 n1 p105-137.

Savignon, S.J. (1983) *Communicative Competence: Theory and Classroom Practice.* Massachusetts: Addison-Wesley Publishing Company.

Stefani, L.A.J. (1994) Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education.* 19, 69-75.

Sundstroem, A. (2005) Self-assessment of knowledge and abilities: A literature study. *EM*, 54, 1-36., available from http://www.nmd.umu.se/digitalAssets/60/60577_em541.pdf, accessed 4 November, 2017.

Von Elek, T. (1981) *Self-Assessment of Swedish as a Second Language*. Göteborg: University of Göteborg (The Language Teaching Research Centre, Work Paper No. 30).

Wilson, K.M. (1999) Validity of Global Self-ratings of ESL Proficiency based on an FSI/ILR-Referenced Scale. ETS Report No. RR-99-13, ETS Research Report Series. Statistics & Research Division, Princeton, NJ.