Assessment & Evaluation in Higher Education

Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners

İhsan Ünaldı[a]

[a] Foreign Languages Teaching Department, Gaziantep University,
Gaziantep, Turkey
Published online: 14 Nov 2014.

CrossMark

Click for updates

PLEASE SCROLL DOWN FOR ARTICLE

Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners

İhsan Ünaldı*

Foreign Languages Teaching Department, Gaziantep University, Gaziantep, Turkey

Although self-assessment of foreign language skills is not a new topic, it has not yet been widely explored in the Turkish English as a Foreign Language (EFL) context. The current study investigates the potential of self-assessment of foreign language skills in determining proficiency levels of Turkish learners of EFL: 239 learners participated in the study. Their receptive language skills were tested with an objective placement test, and the results were compared with the grades assigned by their instructors and the results of criterion-referenced self-assessment scores. Multiple regression analysis revealed that teacher and self-assessment scores were significantly correlated with each other; however, teacher assessment, compared to self-assessment, appeared to be a much stronger predictor of the actual proficiency levels of the participants. In addition to this, participants at lower level of proficiency revealed a common tendency of overestimating their language skills, while with the higher level learners there were clear signs of underestimation.

Keywords: self-assessment; teacher assessment; foreign language proficiency

Introduction

There has been a considerable shift from teacher-centred to learner-centred instruction, especially in foreign language pedagogy (Anton 1999; Harmer 2001; Nunan 2013). Today, it is virtually impossible to find serious support for language instruction depending totally on the teacher. This view has been shaping language pedagogy for decades now. The perspectives of the Common European Framework of Reference for Languages (CEFR) could be counted as one result, or maybe the cause at some levels. Self-assessment, although not totally new, is one of the key concepts for discussion.

In Turkey, there has been official emphasis on and support for certain concepts provided by the CEFR (Mirici 2008). The CEFR standards are very likely to shape language pedagogy in Turkey in the long run. However, classroom observations, teachers' room discussions and anecdotal evidence suggest that concepts like learner autonomy and related issues, which are repeatedly emphasised in CEFR, do not seem to have much support in the Turkish educational context. One of these issues, self-assessment in higher level education in relation to language proficiency and formal assessment, is the focus of this study.

Within the framework of higher education, 'the self-theories of students are of obvious significance' (Yorke and Knight 2004, 29). The results of this study will be

*Email: unaldi@gantep.edu.tr

important, especially in its own context, because, if the integration of self-assessment of foreign language skills into the higher education system in Turkey is a real concern, we need to know whether Turkish English as a foreign language (EFL) learners are able to make sound judgements about their own language proficiency levels.

The merits and importance of getting students to self-assess might change from one context to another. There are some basic differences among young learners, teenagers and adult learners in pedagogical terms. Adult learners are known to be more self-directing, experienced and internally motivated (Knowles 1980). Assessing young learners or teenagers based on the teacher's criteria might sound logical, as this group of learners tend to need approval from the teacher (Harmer 2001). Adult learners need to be assessed by taking into account their pedagogical differences. The main purpose of the current study is to deal with this particular aspect of adult learning. If it can be shown that language learners in higher education can assess their language proficiency levels to a certain extent, this will have practical implications.

Overview of the related literature

The use of the term self-assessment goes back as far as the 1980s. Dickinson (1982), for example, considers self-assessment as an aspect of learner autonomy, Blanche (1988) sees it as a condition, and Harris (1997) regards it as one of the pillars of learner autonomy. If students can evaluate their own performance to a certain extent, then their teachers will become aware of their learning needs (Harris 1997). It is also noted that, although it could become a part of it, self-assessment is not self-grading (LeBlanc and Painchaud 1985). The Council of Europe (2001) briefly defines self-assessment as 'judgments about your own proficiency' (191). The focus of this definition is obviously on the learner, not on some outside official authority who uses standardised test results to reach such judgements.

Over the years, self-assessment has been the subject of various studies from different perspectives. These include the comparison of self-assessment with peer assessment (Boud and Falchikov 1989; Patri 2002; Matsuno 2009), its effects on young learners (Goto Butler and Lee 2010), its relationship with learning styles (Cassidy 2007), its relation to teacher assessment (Matsuno 2009; Leach 2012; López-Pastor et al. 2012), and its use in foreign language learning (LeBlanc and Painchaud 1985; Blanche 1988; Blue 1994; Hargan 1994; Deville and Deville 2003; De Leger 2009; Goto Butler and Lee 2010; Crusan 2011; Brantmeier, Vanderplank, and Strube 2012).

In some of these studies, evidence of a tendency for learners to underestimate their performance has been found. For example, among the four approaches to learning (deep, surface, strategic and apathetic), surface learners appeared to be inclined to provide lower evaluations of their own performance (Cassidy 2007). Leach (2012) analysed 120 student responses to self-assessment and found that higher achieving students tended to underrate and lower achieving ones were more inclined to overrate. However, it was suggested that this might be a function of the grading scale rather than necessarily reflecting students' capability to self-assess.

In the related literature, there are numerous studies with particular reference to EFL settings. In a meta-analysis study, Ross (1998) states that the largest number of correlations is between second language reading variables and self-assessments in

reading, one of the receptive skills. Under certain conditions, self-assessment is seen as a powerful instrument in second language placement (LeBlanc and Painchaud 1985). Comparisons of multiple-choice grammar-based and self-placement tests are also discussed in the literature, for example, Hargan (1994) reports the results of a do-it-yourself placement test, comparing it with a multiple-choice test. It is claimed that these two tests yielded quite similar results. Similarly, Deville and Deville (2003) regard self-assessment as another possibility to obtain an individual's initial second language ability and use their estimates as a starting point.

Goto Butler and Lee (2010) examined the effectiveness of self-assessment among 254 Korean 6th grade EFL learners. The participants were asked to perform self-assessment on a regular basis, and the analyses revealed positive effects on the participants' performance and confidence in learning English. De Leger (2009) researched the evolution of language learner's perception of themselves as second language speakers, of French in this case, through the use of self-assessment tools. It was observed that their self-perception evolved positively over time in terms of fluency, vocabulary and overall confidence in speaking.

Self-assessment is based on judgements about one's own foreign language proficiency level; however, one could argue that such judgements are beyond reach of most language learners. From this perspective, doubts have been cast on the reliability of self-assessment in formal education because of its serious limitations (Blue 1988, 100), or the inability of language learners to evaluate themselves due to lack of training (Jafarpur 1991). These arguments might put self-judgements of language learners in jeopardy. Some research findings failed to show a significant difference between self and teacher assessment (Leach 2012) in domains other than language pedagogy; however, when EFL learners were compared in terms of teacher evaluation, self-assessment and peer-rating, the results indicated that self-assessment scores of the subjects who had had no training in evaluation did not seem to be correlated with teacher assessment (Jafarpur 1991). Yorke and Knight (2004) argue that the interaction between the self-theories held by teachers and students may be particularly significant. Another insight about this issue is that teacher assessments could be best used when the criterion variables are related to general proficiency, rather than specific course objectives (Ross 1998).

It has also been reported that many college level students in English for academic purposes classrooms were not influenced by their scores on internationally recognised language tests when they assessed their own language level (Blue 1994). After comparing the international language test sores of the participants with their self-ratings of language ability, the researcher claims that these ratings are not reliable. For the learners, self-assessment seems to be a very difficult task even when teacher feedback is involved in the process. However, the claim that language learners who have realistic ideas about their language proficiency levels are more likely to continue with language learning, when compared to those with unrealistic ideas about the issue, seems to be supported by the findings.

Related research with English as second language learners which takes into account second language proficiency levels is quite limited. In one of these studies it was revealed that advanced learners underestimate their second language abilities (Kathy Heilenman 1990). Only a weak correlation between self-assessment of language proficiency and objective measures of language proficiency is mentioned in the literature (Peirce, Swain, and Hart 1993), and this weak correlation is used to question the reliability of SA; however, no account is given as to the correlation

between these two variables across different proficiency levels. The Council of Europe (2001) mentions that 'at least adult learners are capable of making such qualitative judgments about their competence' (191), but again proficiency levels are not mentioned either.

In the studies that have been mentioned so far there seem to be conflicts as to whether learners have the capacity to make sound assessments about their own learning. It seems that there is a need for the identification of learners who can assess themselves if self-assessment is to be a part of higher education. The lack of comparisons between the self-assessment scores of EFL learners from different proficiency levels is an issue here. One recent study examines Spanish language learners' skill-based self-assessment scores across three proficiency levels – beginning, intermediate and advanced – and the findings seem to be validating the relationship between self-assessment and achievement tests with advanced learners (Brantmeier, Vanderplank, and Strube 2012).

Apart from the studies that have been mentioned so far, as is usually the case, the research questions of the current study primarily stemmed from recurrent observations. When the topic of discussion among Turkish EFL learners is language proficiency levels, it has been observed that, compared to their general academic outcomes, they either underestimate their proficiency levels or sometimes come up with unrealistic overestimations. These recurrent observations, along with the results of the related studies, underlie the following two research questions:

Research question 1: To what extent can self and teacher assessment scores predict language proficiency levels of Turkish EFL learners?

Research question 2: Do self-assessment accuracy levels of Turkish EFL learners change significantly according to their English language proficiency levels?

Method

Participants and sampling

The total population of the current study consists of 850 freshman students at Gaziantep University, Turkey. The students at this university, unless they succeed in the exemption test which is given at the beginning of the academic year, go through an intense English language programme for a year in the Higher School of Foreign Languages at the university. The learners at this institution are assessed by taking about 20 weekly, 5 monthly and a final test. The tests are prepared by a testing office, which makes the process as objective and accurate as possible. However, the assessment process predominantly focuses on receptive language skills, i.e. reading and listening. In addition to these formal summative grades, the instructors evaluate the learners' overall performances, mainly for receptive skills, and these grades are taken into account while calculating the passing grades. In the current study, only the grades which were assigned by the instructors are considered.

Mainly for practical reasons, cluster sampling was chosen as the sampling method. This method is useful 'especially when the target population is widely dispersed', and the aim 'is to randomly select some larger groupings or units of the populations (for example, schools) and then examine all the students in those selected units' (Dörnyei 2007, 98). Accordingly, 8 freshman groups out of 31 were chosen randomly and a total of 250 students participated in the current study.

Data collection tools

Three tools were used to gather data from the participants. The first one is a simple learner profile form, which involves inputs about the participants like student ID, gender and age. This form also includes a section in which the participants provide permission to the researcher to use the data gathered for research.

The second tool which was used to gather data was a proficiency test developed by Allen (1992). This is a skill-based proficiency test whose validity and reliability has been proved many times. It consists of two main parts with 200 questions. The first part tests listening skill, and the second part is related to grammar and reading. The test was designed to test receptive language skills, and it provides a relatively more economical and easier way of testing proficiency level in English. A score equivalency chart also provides a general guide to the instructor as to the test-taker's place on a number of international scales like IELTS, TOEFL and CEFR.

The third tool used for data collection is a criterion-referenced self-assessment checklist, including can-do statements which were taken from the CEFR. These statements, which were developed for the DIALANG project, are organised to check the four main skills, listening, speaking, reading and writing; however, the self-assessment checklist that was used in the current study included only the receptive skills, listening and reading, in order to match it with the proficiency test and teacher assessment. DIALANG is a project which aims at developing 'an assessment system intended for language learners who want to obtain diagnostic information about their proficiency' (Council of Europe 2001, 226). It consists of self-assessment, language tests and feedback for 14 European languages. In this system descriptive scales, which are based on the CEFR, are used as diagnostic tools for self-assessment. One of these tools is the checklist: this includes 74 can-do statements, and the participant gets one point for each positive response. Therefore, it is assumed that the more positive answers a participant marks, the higher his/her self-assessment score will be. The main reason for choosing this tool is that 'learners will be more accurate in the self-assessment process if the criterion variable is one that exemplifies achievement of functional ("can do") skills on the self-assessment battery' (Ross 1998, 16).

The pilot study and data collection procedures

Before the initiation of the procedures, the research tools were piloted with a small group of students ($n = 28$). In order to test the reliability of self-assessment tool, it was applied to the same group after two weeks, and the two scores of the individuals were analysed for correlation. The analysis yielded a significant correlation ($r = .603$; $n = 28$; $p = .001$). For validity concerns, the first scores of the self-assessment tool were tested for correlation with the proficiency test, and this analysis also revealed a significant correlation between the two sets of scores ($r = .511$; $n = 28$; $p = .006$).

For ethical concerns, official permissions were obtained from the related institution. In the first sessions with the groups, the participants were handed out the learner profile forms and consent letters to be signed. In the following sessions with the groups, which were a week later, they were given the placement test. The testing sessions lasted for about 90 min for each group. In the third sessions, the participants were asked to assess themselves by using the self-assessment tool, and this process took about 40 min for them to complete.

In order to increase the overall accuracy of the data, all of the groups were given clear explanations concerning the process: sometimes further explanations were needed in some groups, and sometimes help was provided to individuals. However, during the process some complications did occur. Some of the participants who took part in the self-assessment activity and took the placement test did not want their results to be used in this study, and they had to be removed from the sampling group. Some other participants who failed to provide required personal information were also removed from the group, and, after these modifications, the total number of the participants reduced to 239 (female $n = 43$; male $n = 196$).

Data analysis

The English placement test consists of 200 items, half of which relates to listening and the other half to reading and grammar. The grading scale provided with the test is very specific, and the test-taker gets one point for each correct answer and the total grade is calculated accordingly. The self-assessment checklist has 74 items in it and the participant gets one point for each positive response. The participants' responses were recorded into an Excel worksheet and proficiency levels were calculated as pre-intermediate, intermediate and upper-intermediate by taking into account the grading scale. Afterwards, the raw data were input into IBM SPSS Statistics (version 21) software package for the analysis process.

To answer the first research question, a multiple regression analysis was carried out. A robust regression analysis is not as straightforward as it might seem. There are strict assumptions that have to be met to be able to make sound interpretations based on the data at hand. In the current study, before beginning the analysis process, these assumptions were tested. First of all, the assumption related to the sampling size was analysed. For a standard regression, at least 20 times more cases than predictors are needed (Coakes 2005, 169). In our case, since there are two predictors, this assumption is met with a sample size of 239.

Next, the issue concerning outliers was tested. With this orientation, Mahalanobis distance values were calculated for each case and no value greater than 10.75 was detected. As is generally accepted, 'in small samples ($n = 100$) and with fewer predictors values greater than 15, and in very small samples ($n = 30$) with only 2 predictors values greater than 11' are potentially problematic (Field 2009, 218). In our case, with a relatively greater sample size ($n = 239$), no outliers were detected.

Another important assumption, multicollinearity, is related to correlations among the independent variables. The existence of too much correlation actually poses serious problems. This assumption can be tested by checking the variance inflation factor (VIF) scores among the coefficients: VIF scores greater than 10 and an average VIF score greater than 1 are serious concerns. These scores in the data-set were checked and no multicollinearity was detected.

Other assumptions for regression analysis like normality, linearity, homoscedasticity and independence of residuals were also examined through residual scatterplots (Coakes 2005, 169). An analysis of the related histogram, scatterplot and P–P plot revealed that the differences between the obtained and predicted proficiency scores were normally distributed, and the residuals were also in a linear relationship with it.

In order to answer the second research question, the results of the participants' proficiency level test and self-assessment scores were used. The results of the

proficiency test revealed that the main population consisted of three groups with different proficiency levels: pre-intermediate, intermediate and upper-intermediate. The related research question was addressed based on this outcome. To be able to make certain comparisons between these two variables with different total scores, the participants' raw scores were first standardised into z-scores and then, to get rid of the negative values, converted into t-scores. In order to see the accuracy levels in self-assessment scores, the standardised self-assessment scores were subtracted from the standardised proficiency scores. With an aim to detect different accuracy levels among the participants, these scores were analysed through a two-step cluster analysis, and the participants were grouped into three by the software according to the differences between their self-assessment and language proficiency scores.

Results

To answer the first research question concerning self and teacher formal assessment scores as predictors of language proficiency levels, a regression analysis was carried out, and the model summary is presented in Table 1.

According to the model summary, participants' proficiency levels significantly correlate with self and teacher assessment ($R^2 = .325$, $F_{(22, 36)} = 56.75$, $p < .05$). It could be claimed that self-assessment, together with teacher assessment, is a strong predictor of proficiency level. However, this correlation does not help us to understand the predictive strengths of self-assessment and teacher assessment individually. In order to determine this difference, the results of the regression analysis can be examined in Table 2.

As can be seen in Table 2, both self-assessment and teacher assessment are significant predictors of participants' actual proficiency levels. However, with larger t and β values, teacher assessment ($\beta = .48$, $p < .05$) seems to be a much stronger predictive when compared to self-assessment ($\beta = .16$, $p < .05$).

Table 1. R^2 table for self-assessment and teacher assessment as the predictor of proficiency level.

Model	R	R^2	Adjusted R^2	Standard error of the estimate
1	.570	.325	.319	10.125

Note: Predictors: (constant), teacher assessment, self-assessment.

Table 2. The results of regression analysis for self-assessment and teacher assessment.

Model	Unstandardised coefficients		Standardised coefficient (β)	t	Significance
	β	Standard error			
1 (Constant)	65.23	5.92		11.01	.000
Self-assessment	.19	.07	.16	2.65	.009
Teacher assessment	.76	.09	.48	8.16	.000

Note: Dependent variable: proficiency test.

Table 3. Proficiency level distribution.

Proficiency levels	f	$\%$
Pre-intermediate	63	26.36
Intermediate	76	31.80
Upper-intermediate	100	41.84
Total	239	100

In order to address the second research question concerning participants' self-assessment accuracy and foreign language proficiency levels, their proficiency level distribution was analysed first. Table 3 provides this distribution.

As can be seen in Table 3, the group consists of mainly three proficiency levels, and the predominant group is the upper-intermediate one with 41.84%. In order to answer the related research question, the scores obtained by subtracting participants' standardised self-assessment scores from their standardised proficiency test scores were processed through a two-step cluster analysis. The results of the cluster formation (ratio of sizes = 2.35) yielded three different levels of self-assessment accuracy among the participants. The assumption was that the first group with positive scores underestimated their own proficiency levels, and the second group's estimations were relatively accurate. Negative scores in the third group denoted overestimation. A clearer comparison can be made by examining Table 4.

It is clear from Table 4 that the groups differ in terms of minimum and maximum scores. These ranges show the difference between the participants' proficiency test and self-assessment scores. Group 1 has positive scores ranging from 3.51 to 34.28, with a mean of 10.95. This means that the individuals in this group underestimate their actual proficiency levels. On the other hand, the individuals in the second group have both negative and positive scores, ranging from −10.00 to 3.04. This group, compared with the other two, has a narrower score range and the scores are relatively closer to zero with a mean of −3.19, which means that the differences between the proficiency test and self-assessment scores are smaller; therefore, this group has relatively more accurate estimations about their actual proficiency levels than the other two groups. The individuals in the third group, with minus scores ranging from −32.15 to −10.48, and a mean of −16.69, seem to be overestimating their own proficiency levels. The boxplot in Figure 1 demonstrates these interpretations.

Figure 1 reveals some useful information about the three groups formed after the two-step cluster analysis. At first glance, the boxplots are comparatively short, which means that the scores within each group are close to each other. The medians for Group 1 (underestimation), Group 2 (accurate estimation) and Group 3 (overestimation) are 9.18, −2.33 and −14.89, respectively. The scores of Group 1 and Group 3 are skewed, whereas Group 2 seems to exhibit quite a normal

Table 4. Comparison of the three groups formed after the cluster analysis.

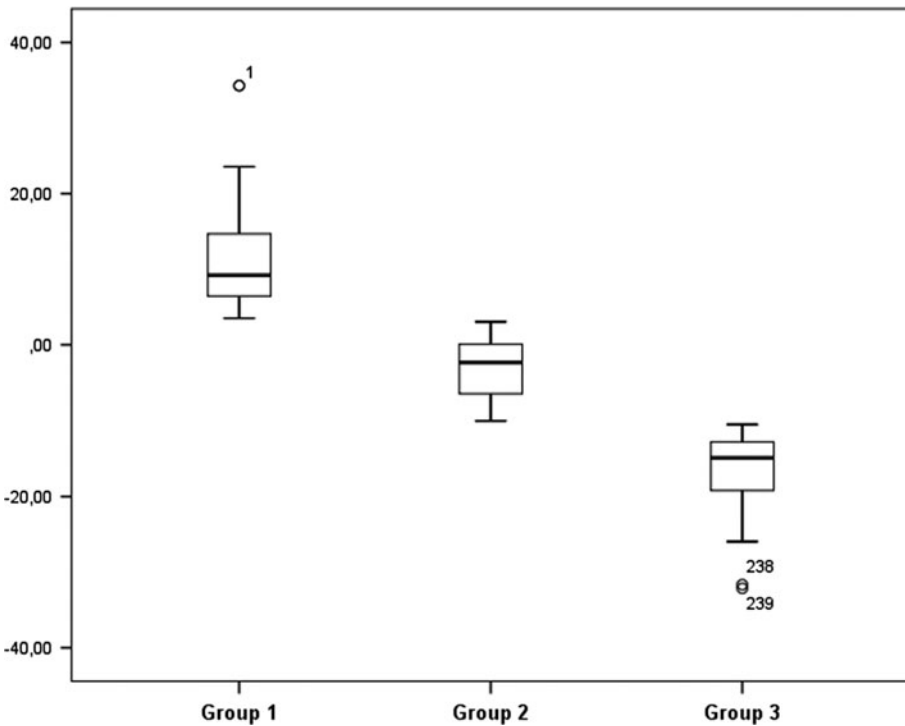|	N	$\%$	Minimum	Maximum	Mean	sd
Group 1	95	39.7	3.51	34.28	10.95	5.72
Group 2	101	42.3	−10.00	3.04	−3.19	3.80
Group 3	43	18	−32.15	−10.48	−16.69	5.42

Figure 1.   Boxplot representation of the clusters formed according to accuracy levels.

distribution. There appears to be one outlier in Group 1, and two in Group 3. In relation to the second research question, these three groups were compared through a chi-square test by taking into account their actual foreign language proficiency levels.

The results of the chi-square test are presented in Table 5. There seems to be a relationship between low level of language proficiency and overestimation of this level because 46% of the participants at pre-intermediate level seem to fall predominantly in the group with a relatively higher level of overestimation (Group 3). On the other hand, 59% of the participants with upper-intermediate proficiency level obviously underestimated their language skills. Furthermore, 101 of the participants (42%) estimated their proficiency levels relatively accurately, and while 81 of them belong to the intermediate and upper-intermediate groups, only 20 of them belong to the pre-intermediate one. In addition, these differences among groups are statistically

Table 5.   Comparison of self-assessment accuracy levels with actual proficiency levels.

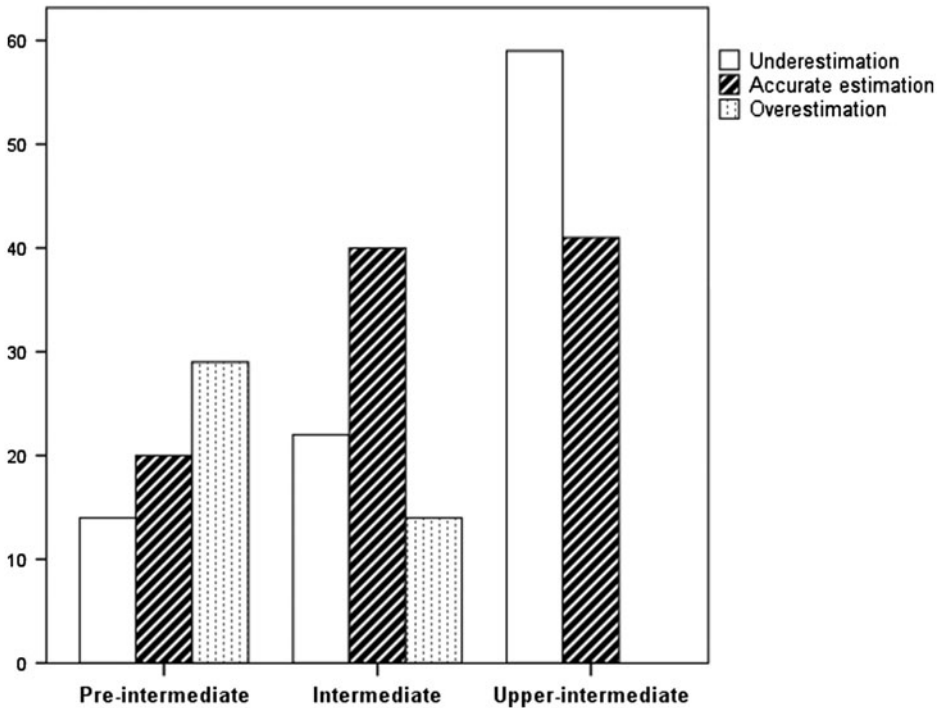|	Group 1		Group 2		Group 3		Total	
	f	%	f	%	f	%	f	%
Pre-intermediate	14	22.22	20	31.75	29	46.03	63	100
Intermediate	22	28.95	40	52.63	14	18.42	76	100
Upper-intermediate	59	59.00	41	41.00	0	0	100	100
Total	95	39.75	101	42.26	43	17.99	239	100

Note: $X^2 = 65.574$; df = 4; $p < .005$; $v = .370$.

Figure 2.   Self-assessment accuracy levels across language proficiency levels.

significant ($X^2 = 65.574$, $df = 4$, $p < .05$) with a strong effect size ($v = .370$). In order to better visualise these differences, a clustered bar chart was formed (see Figure 2).

It is clear from Figure 2 that the pre-intermediate group has the lowest self-assessment accuracy level with a very high tendency of overestimation. Intermediate and upper-intermediate groups are quite close in terms of accuracy. The intermediate group has a clear tendency for both underestimation and overestimation. However, the upper-intermediate group reveals no sign of overestimation but rather a very high level of underestimation.

Discussion

The results of the statistical analyses revealed somewhat satisfactory answers to the research questions. In answer to the first research question, both self-assessment and teacher assessment are strong predictions of proficiency levels of Turkish EFL learners, but, when compared to self-assessment, teacher assessment seems to be a much stronger predictor. As for the second research question regarding accuracy level of self-assessment across proficiency levels, EFL learners at low language proficiency level overestimate their foreign language skills; on the other hand, those at high proficiency level seem to have lower opinions about their skills than would be expected, and the focus of the discussion will be on these two issues.

To begin with, the results of the multiple regression analysis revealed that self-assessment and teacher assessment correlate significantly with each other and together they have a potential to make deductions about learners' actual proficiency

levels. This outcome might be interpreted in a variety of ways. From a broader point of view, teacher and self-assessment have been reported to be in a positive correlation (Leach 2012; López-Pastor et al. 2012). In terms of language pedagogy, however, the results of the studies comparing self and teacher assessment appear to denote different results. Jafarpur's (1991) conclusion was that untrained EFL learners' self-assessment scores did not correlate with teacher assessment. For some, compared to peer assessment, teacher assessment appeared to be less consistent (Matsuno 2009). However, within the confines of the current study, teacher assessment proved to be a very important parameter in predicting EFL proficiency level.

On the other hand, the results concerning self-assessment accuracy levels revealed similar patterns with the related literature. For example, a weak correlation between self-assessment and objective measures of language proficiency is mentioned (Peirce, Swain, and Hart 1993). The results of the regression analysis of the current study support this finding. Although self-assessment is an important predictor of language proficiency, it seems to be much weaker than teacher assessment. The reason for this might be related to the differences of self-assessment accuracy levels across different proficiency levels. The underestimation of foreign language skills by more proficient learners is also in line with the prior research (Boud and Falchikov 1989; Kathy Heilenman 1990; Matsuno 2009). In a different domain than foreign language pedagogy, Leach (2012) reports both underrating and overrating, with slightly more students underrating (37%) than overrating (32%). In our case, underestimation of language proficiency level seems to be at a much higher rate with 40% when compared to overestimation with a rate of 18% (see Table 4). When this situation was analysed through a chi-square test, by taking into account the results of an objective proficiency test, more proficient learners appeared to be underestimating their language skills.

The reasons for this kind of underestimation might be associated with sociocultural issues (Matsuno 2009). Since in many eastern societies modesty is encouraged, some of the more proficient Turkish EFL learners, being aware of their success, might be giving signs of modesty. Whatever the reason might be, underestimation of language skills will pose problems if self-assessment is going to become a part of EFL assessment. The same group, along with the intermediate one, has a significantly higher self-assessment accuracy level compared to the lower proficiency group (see Figure 2). The pre-intermediate group, with some level of accuracy, reveals signs of both underestimation and overestimation of language skills; in addition, this group has the highest level of overestimation. These divergences are probably the reasons for the lower predictive strength of self-assessment compared to teacher assessment.

At the planning stage of the current study, the researcher's intention was to compare different proficiency levels by carrying out regression analysis for each level individually. Three different proficiency levels appeared after the proficiency test, and regression analyses were carried out for each level. However, when the sampling group was further divided into three different groups, nearly all of the assumptions for a sound regression analysis were violated. The researcher acknowledges that this situation could have been avoided if the sampling size had been greater.

The students who participated in this study might be overestimating or underestimating their language skills simply because of the grading scale. As Leach (2012) points out, high-level learners are more likely to overrate and low-level ones will probably underrate because they have more opportunities to do so. These two

groups could be regarded as the far ends of the spectrum; therefore, if potential problems concerning self-assessment in language pedagogy are to be addressed, more attention should be paid to the intermediate group. The overestimation and underestimation of language skills among EFL learners are real issues; however, the self-assessment accuracy level of the intermediate group is high, which means that there is room for development.

The points made so far indicate that, contrary to anecdotal evidence, the Turkish EFL context in higher education might benefit from self-assessment. It has been claimed that language learners with realistic ideas about their language proficiency levels have greater chances to continue with their learning when they are compared to those with unrealistic ideas about the same issue (Blue 1994). In addition, self-assessment could be used to help learners to build their confidence in learning English, thus positively affecting their overall performance (De Leger 2009; Goto Butler and Lee 2010).

It has also been reported that self-assessment could be an effective complement to formal teacher assessment unless it is a part of high-stakes testing (Council of Europe 2001, 91). If the learners are trained about self-assessment, higher correlations between teacher and self-assessment are also possible. The participants in the current study, albeit not very precisely, were able to assess their own language skills to an extent that the scores that they assigned for themselves are significantly correlated with the results of an objective proficiency test and the scores that their instructors assigned. Considering that this correlation exists without any self-assessment training on the learners' side, it seems that, as is claimed by McDonald and Boud (2003), there is a great deal of potential for self-assessment scores to become an important and effective part of formal assessment.

The results of the current study are limited to receptive foreign language skills; therefore, it seems necessary to carry out a similar study of the productive skills of Turkish EFL learners. In addition, in the current study the participants were first assessed by their instructors and then assessed themselves after a period of time. It could be argued that this could have shaped participants' self-perceptions in terms of language skills. The possible effects of teacher assessment over self-assessment could be the focus of another study.

Conclusion

The present study reveals the potential of teacher and criterion-referenced self-assessment in predicting the proficiency levels of EFL learners. However, it should be noted that this potential seems to get weaker at lower proficiency levels. If a learner-friendly environment is one of the goals of a language teaching programme, self-assessment should be included in the process to increase learners' motivation and help avoid the negative aspects of formal assessment. While doing so, as the results of the current study indicate, some caution is needed with learners at lower proficiency levels.

Notes on contributor

İhsan Ünaldı is a full-time assistant professor in the English Language Teaching department at Gaziantep University. Prior to joining the academics, he taught English to a variety of age and level groups. Although his main area of interest is corpus linguistics, he is also carrying out studies on second language vocabulary instruction, assessment and English literature.

References

Allen, D. 1992. *Oxford Placement Test 2*. Oxford: Oxford University Press.

Anton, M. 1999. "The Discourse of a Learner-centered Classroom: Sociocultural Perspectives on Teacher-learner Interaction in the Second-language Classroom." *The Modern Language Journal* 83 (3): 303–318. doi:10.1111/0026-7902.00024.

Blanche, P. 1988. "Self-assessment of Foreign Language Skills: Implications for Teachers and Researchers." *RELC Journal* 19 (1): 75–93. doi:10.1177/003368828801900105.

Blue, G. 1988. "Self-assessment-the Limits of Learner Independence." In *Individualisation and Autonomy in Language Learning*, edited by A. Brookes, 100–118. London: Modern English Publications/British Council.

Blue, G. 1994. "Self-assessment of Foreign Language Skills: Does It Work?" *CLE Working Papers* 3: 18–35.

Boud, D., and N. Falchikov. 1989. "Quantitative Studies of Student Self-assessment in Higher Education: A Critical Analysis of Findings." *Higher Education* 18 (5): 529–549.

Brantmeier, C., R. Vanderplank, and M. Strube. 2012. "What about Me?" *System* 40 (1): 144–160. doi:10.1016/j.system.2012.01.003.

Cassidy, S. 2007. "Assessing 'Inexperienced' Students' Ability to Self-assess: Exploring Links with Learning Style and Academic Personal Control." *Assessment & Evaluation in Higher Education* 32 (3): 313–330. doi:10.1080/02602930600896704.

Coakes, S. 2005. *SPSS: Analysis without Anguish Using SPSS: V. 12*. 1st ed. Milton: Wiley.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. http://www.coe.int.

Crusan, D. 2011. "The Promise of Directed Self-placement for Second Language Writers." *TESOL Quarterly* 45 (4): 774–774. doi:10.5054/tq.2010.272524.

De Leger, D. S. 2009. "Self-assessment of Speaking Skills and Participation in a Foreign Language Class." *Foreign Language Annals* 42 (1): 158–178.

Deville, M., and C. Deville. 2003. "Computer Adaptive Testing in Second Language Contexts." *Annual Review of Applied Linguistics* 19 (Aug.): 273–299. doi:10.1017/S0267190599190147.

Dickinson, L. 1982. *Self-assessment as an Aspect of Autonomy*. Edinburgh: Scottish Centre for Overseas Education.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies*. Oxford: Oxford University Press.

Field, A. 2009. *Discovering Statistics Using SPSS: And Sex and Drugs and Rock 'N' Roll*. 3rd ed. London: Sage.

Goto Butler, Y., and J. Lee. 2010. "The Effects of Self-assessment Among Young Learners of English." *Language Testing* 27 (1): 5–31. doi:10.1177/0265532209346370.

Hargan, N. 1994. "Learner Autonomy by Remote Control." *System* 22 (4): 455–462. doi:10.1016/0346-251X(94)90002-7.

Harmer, J. 2001. *The Practice of English Language Teaching*. London: Longman.

Harris, M. 1997. "Self-assessment of Language Learning in Formal Settings." *ELT Journal* 51 (1): 12–20. doi:10.1093/elt/51.1.12.

Jafarpur, A. 1991. "Can Naive EFL Learners Estimate Their Own Proficiency?" *Evaluation & Research in Education* 5 (3): 145–157.

Kathy Heilenman, L. 1990. "Self-assessment of Second Language Ability: The Role of Response Effects." *Language Testing* 7: 174–201.

Knowles, M. 1980. *The Modern Practice of Adult Education: From Pedagogy to Andragogy*. New York: Cambridge Books.

Leach, L. 2012. "Optional Self-assessment: Some Tensions and Dilemmas." *Assessment & Evaluation in Higher Education* 37 (2): 137–147. doi:10.1080/02602938.2010.515013.

LeBlanc, R., and G Painchaud. 1985. "Self-assessment as a Second Language Placement Instrument." *TESOL Quarterly* 19 (4): 673–687.

López-Pastor, V. M., Juan-Miguel Fernández-Balboa, María L. Santos Pastor, and Antonio Fraile Aranda. 2012. "Students' Self-grading, Professor's Grading and Negotiated Final Grading at Three University Programmes: Analysis of Reliability and Grade Difference Ranges and Tendencies." *Assessment & Evaluation in Higher Education* 37 (4): 453–464. doi:10.1080/02602938.2010.545868.

Matsuno, S. 2009. "Self-, Peer-, and Teacher-assessments in Japanese University EFL Writing Classrooms." *Language Testing* 26 (1): 075–100. doi:10.1177/0265532208097337.

McDonald, B., and D. Boud. 2003. "The Impact of Self-assessment on Achievement: The Effects of Self-assessment Training on Performance in External Examinations." *Assessment in Education: Principles, Policy & Practice* 10 (2): 209–220. doi:10.1080/0969594032000121289.

Mirici, İ. H. 2008. "Development and Validation Process of a European Language Portfolio Model for Young Learners." *Turkish Online Journal of Distance Education TOJDA* 9 (2): 26–34.

Nunan, D. 2013. *Learner-centered English Language Education: The Selected Works of David Nunan*. New York: Taylor and Francis.

Patri, M. 2002. "The Influence of Peer Feedback on Self and Peer-assessment of Oral Skills." *Language Testing* 19 (2): 109–131.

Peirce, B. N., M. Swain, and D. Hart. 1993. "Self-assessment, French Immersion, and Locus of Control." *Applied Linguistics* 14 (1): 25–42. doi:10.1093/applin/14.1.25.

Ross, S. 1998. "Self-assessment in Second Language Testing: A Meta-analysis and Analysis of Experiential Factors." *Language Testing* 15 (1): 1–20. doi:10.1177/026553229801500101.

Yorke, M., and P. Knight. 2004. "Self-Theories: Some Implications for Teaching and Learning in Higher Education." *Studies in Higher Education* 29 (1): 25–37. doi:10.1080/1234567032000164859.