

What about me? Individual self-assessment by skill and level of language instruction

Cindy Brantmeier ^{a,*}, Robert Vanderplank ^{b,1}, Michael Strube ^{c,2}

^a *Department of Romance Languages & Literatures, Washington University in St. Louis, Campus Box 1077, One Brookings Drive, St. Louis, MO 63130-4899, United States*

^b *Oxford University Language Center, 12 Woodstock Road, Oxford, England OX2 6HT, United Kingdom*

^c *Department of Psychology, Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899, United States*

Received 3 May 2011; revised 18 November 2011; accepted 13 December 2011

Abstract

In an investigation with advanced language learners, Brantmeier [Brantmeier, C., 2006. Advanced L2 learners and reading placement: self-assessment, computer based testing, and subsequent performance. *System* 34 (1), 15–35.] reports that self-assessment (SA) of second language (L2) reading ability, when measured with self-rated scales, is not an accurate predictor of subsequent reading performance as measured via multiple choice items. In another experiment with advanced learners that utilizes criterion-referenced SA items, Brantmeier and Vanderplank [Brantmeier, C., Vanderplank, R., 2008. Descriptive and criterion-referenced self assessment with L2 readers. *System* 36 (3), 456–477] reveal that learners accurately estimate their reading comprehension when it is measured via multiple choice items. For the present study, an SA instrument of language learning achievement was designed according to specific course content to take into consideration the direct experience learners have had in practicing reading, listening, speaking, and writing [Ross, S., 1998. Self-assessment in second language testing: a meta-analysis of experimental factors. *Language Testing* 15, 1–20.]. With 276 participants, the study examines skill-based SA across beginning, intermediate and advanced levels of language instruction, and it offers evidence to validate the relationship between the SA instrument and achievement on an online abilities test with advanced learners. Findings hold important implications for language learner assessment, especially in terms of advanced students' ability to rate themselves when given specific criteria. A discussion about the value of SA as a complement to other traditional approaches for language program assessment is offered.

© 2012 Elsevier Ltd. All rights reserved.

Keywords: Advanced, intermediate, and beginning language learners; Self-assessment; Language program assessment

* Corresponding author. Tel.: +1 314 935 7953; fax: +1 314 726 3494.

E-mail addresses: cbrantme@wustl.edu (C. Brantmeier), robert.vanderplank@lang.ox.ac.uk (R. Vanderplank), mjstrube@wustl.edu (M. Strube).

¹ Tel.: +44 1865 2 83363.

² Tel.: +1 314 935 6545.

1. Introduction

The main goal for university assessments of second and foreign languages is to provide evidence that performance in a specific curriculum is changing and improving over time. As part of this overall picture, assessment of L2 learner's ability to read, write, listen, and speak across stages of acquisition has become of vital importance at most universities. The evaluation of learner abilities upon completion of each stage serves as a basis for substantiating whether or not learners are improving their skills as they advance to higher levels of language learning.

At some American universities, language departments create their own summative tests for each level of instruction that become part of a portfolio of assessment. The tests are often completed during class time and do not affect course grades. There appears to be a great deal of scepticism from language program directors and course coordinators about summative tests that are used solely for external evaluative purposes. The learner may experience a great deal of anxiety and stress while completing what they may view as a high-stakes exam. Consistent evaluation and testing of all four skills throughout the semester is already a part of many language courses, and the distribution of factors for final course grades further reveals the process of learning across different skills. Consequently, results of course exams, which are driven by the content of course syllabi, along with course grades may be part of the language assessment portfolio. Additionally, SA instruments may provide rich evidence about individual and collective achievement as students evaluate their own learning and consolidate their language learning experiences. The purpose of the present study is to examine whether skill-based SA items provide confirmation about the nature and quality of language learning with different skills across levels of instruction.

2. Literature review — self-assessment of L2 abilities

Blanche and Merino (1989) detailed studies that examine SA in language testing and concluded that the accuracy of SA depends on the linguistic skills being assessed as well as the materials involved in the testing situation. They reported that overall learner SA estimates are good or very good (p. 315). Some years later, later, Ross (1998) provided a detailed meta-analysis of research that gives a cumulative depiction of investigations in this area and demonstrated that there is variation in how accurately learners assess their own second language skills. Ross offered a thorough discussion that substantiates the need to consider direct experience and specific curricular content as factors when designing SA instruments. In Ross's terms, direct experience refers to the learning and practice learners have had in all four skills. Much prior research on SA and language learning includes participants from the beginning stages of instruction instead of the advanced stages (Bachman and Palmer, 1981, 1982, 1996; Buck, 1992; Ferguson, 1989; LeBlanc and Painchaud, 1985; Wongsotorn, 1981).

Table 1 lists selected studies that examine SA and language learning from 1978 to 2010. The table highlights the skills addressed in each study along with results. As outlined in the table, most of the earlier work utilized adult learners of English and emphasized listening and reading. The findings of the studies are inconclusive, which is not surprising given the great deal of variation and disparities in the design and content of SA items and inventories as well as in the skills being tested. For example, Ross (1998) found that students more accurately assessed themselves with an SA inventory that contained questions about functional skills as opposed to one that includes SA items of abstract nature. In a later study, Brantmeier (2005) examined SA of L2 reading abilities and utilized a scale for self-rating. More specifically, students were asked to rate themselves as readers of Spanish. Possible answers included 5 options from "I think I am an excellent reader of Spanish" to "I do not think I am a good reader of Spanish at all." Findings held positive for SA and reading comprehension with recall tasks but not with multiple choice tasks. Subsequently, Brantmeier (2006) used a contextualized inventory after students read in Spanish along with a set of self-rating items. Her findings were at variance with prior research and indicated no positive relationship with all SA items with three different reading comprehension tasks. With students of English, Butler and Lee (2006) used on-task formats for SA questions that consisted of items completed directly after completing specific tasks, and reported that this rubric was more accurate than off-task formats, where students did not actually attend to the language.

In a more recent experiment with advanced learners of Spanish that utilized criterion-referenced SA items, or items that included a description of the specific knowledge and skills each student demonstrates, Brantmeier and Vanderplank (2008) found that learners were able to accurately estimate their reading comprehension as measured via multiple choice items. In this experiment, researchers created SA items that were grounded in self-assessment criteria established by the DIALANG project (<http://www.lancs.ac.uk/researchenterprise/dialang/about>) a testing

Table 1
Selected studies on SA and language learning abilities.

Author/year	Participants	Skill addressed	Results
Oskarsson (1978)	Adult students of EFL in Sweden	Reading and listening	SA positively correlates with teacher ratings and written test scores
LeBlanc and Painchaud (1985)	First-year students of English and French at a university in Canada	Listening and reading	Positive correlations between SA and standardized proficiency tests
Krausert (1991)	Adult students of ESL	Reading, writing and speaking	Positive relationship between SA of all three abilities
Hargan (1994)	Adult students of ESL		A multiple choice test and SA test indicate same level of language placement
Birckbichler et al. (1993)	Adult students of ESL in the USA	Reading and writing	SA correlated higher than any other variable with reading comprehension scores
Ross (1998)	Adult students of English in Japan	Speaking, listening, reading and writing	Learners are more accurate in SA if battery contains functional skills. With SA items of abstract nature, SA is less accurate
Deville and Deville (1999)	Adults of learning different languages in the USA	Speaking, listening, reading and writing	SA is an effective starting point for CAT for placement
Brantmeier (2005)	Adult students of Spanish at university in USA	Reading	SA positively correlates with enjoyment and written recall, but not with multiple choice items
Brantmeier (2006)	Adult students of Spanish at university in USA	Reading	SA is not an accurate predictor of reading on a computer based test and subsequent reading performance
Butler and Lee (2006)	Elementary school students learning English in Korea	Speaking	SA with on-task format is more accurate than SA with off-task format
Brantmeier and Vanderplank (2008)	Adult students of Spanish at university in USA	Reading	Learners do accurately SA reading when comprehension is measured via multiple choice
De Saint-Leger (2009)	Adult students of French in Australia	Speaking	SA evolves positively over time in relation to fluency, vocabulary and self-confidence
De Saint-Leger and Storch (2009)	Adult students of French in Australia	Speaking	SA affects willingness to communicate in class as well as confidence
Little (2009)	Adults and children learning English in Ireland	Speaking, listening, reading and writing	SA fosters learner autonomy and aids with learner reflection
Wan-a-rom (2010)	High school students learning English in Thailand	Reading	SA directs learners to an appropriate reading level for extensive reading curriculum
Baniabdelrahman (2010)	High school students learning English in Jordan	Reading	SA positively affects reading performance

Updated and modified from Brantmeier and Vanderplank (2008).

system for 14 different European languages (see Alderson, 2005, for a complete history and description of the DIALANG project). In another recent investigation with students studying French, De Saint-Leger (2009) examined how SA abilities with speaking change throughout a semester. She found that students' ability to assess themselves accurately developed over time. In a similar study, De Saint-Leger and Storch (2009) reported that SA positively affected learners' willingness to communicate orally in class and it also helped with issues of self-confidence. In both of these studies, the researchers used SA questionnaires that included multiple choice items, self-rating scales, and open-ended questions. Little (2009) described how criterion-referenced SA items, taken from the European Language Portfolio, stimulated reflective learning with SA playing a central role. Recently, Wan-a-rom (2010) examined the SA of specific lexical knowledge with graded readers and offered positive support for the use of SA in extensive reading programs. Even more recently, for SA measures with Jordanian students of English Baniabdelrahman (2010) utilized one-minute papers, where the students wrote down short answers to two different questions about the most important point made in class that day, and rating scale sheets (a written list of performance criteria associated with a particular activity) and found that self-assessment has a positive effect on students' performance in reading English.

Given the inconsistencies and disparities in instruments used to measure self-assessment, along with the variations of skills and diagnostic tests, generalizations about SA and language learning across stages of acquisition cannot yet be made. What can be said about the SA instrument that was validated and used for the DIALANG project, however, is that students assess their abilities better at higher levels of language instruction than at lower levels (Alderson, 2005). Findings are consistent with the hypothesis that there is a threshold beyond which learners are better able to assess themselves and learn independently. For example, Stern (1983) contends that learners may need to complete the intermediate level before they can really handle independent language learning, as students at the beginning and intermediate levels tend to be quite teacher dependent (p. 353). It is important to note that the criterion-referenced SA items used in the DIALANG project proved reliable with data-driven evidence, and that the DIALANG test is not a high or low-stakes exam as it is for individuals who want to self-diagnose their language abilities. Results from the test are not used for any kind of language program assessment. The present study attempts to examine whether or not the SA component of the DIALANG project could serve as an additional element for language learner assessment in the USA, and it also attempts to validate the SA instruments with advanced language learners, who are beginning the language major, by comparing SA to actual test scores.

3. The present study

Alderson (2005) discusses how SA is the main component of DIALANG, and again, the emphasis of the DIALANG tests is for students to assess their skills in a low-stakes environment. The SA instruments consist of criterion-referenced items where learners situate themselves in a language task and then evaluate their own performances. For the present investigation, items on the DIALANG tests were modified and tailored to meet objectives for each course at different stages of acquisition. Further explanation of the development of the SA items for this study will be given in the next section.

3.1. Methods

The following research questions guide the present study:

1. When measured with a criterion-referenced instrument, what are the self-ratings of reading, writing, listening, and speaking skills with beginner, intermediate, and advanced learners of Spanish?
2. With advanced learners of Spanish, is there a relationship between criterion-referenced self-assessed ratings of skills and subsequent performance?

3.2. Participants

Participants in the present study were 276 students, ages 19–22, enrolled in beginning 1 ($n = 28$), beginning 2 ($n = 37$), intermediate ($n = 38$), or advanced ($n = 150$) level language courses of Spanish at a private university in the Midwest of the United States. The university does not have a language requirement and consequently, enrollment in the courses is by choice. The advanced language courses mark the beginning of the Spanish major.

During the second to the last week of classes, all participants across all levels of Spanish instruction completed a SA instrument during class time, which took approximately 25 min. After this, outside of class, 150 participants who were enrolled in Advanced Spanish (Spanish 307) also completed a skill-based online diagnostic test. The majority of students at the university place directly into this advanced level as incoming freshmen, and therefore the total number of students is greater than the total number of students enrolled in the intermediate and beginning levels. Furthermore, students at this advanced level of language instruction are enrolled in the course that begins the Spanish major. Results of both the SA instrument, utilized with students in all three levels, and the diagnostic test, used with advanced language learners, did not influence course grades or students' ability to enroll in higher level courses. It is important to note that students did not complete a test of this format previously, and that all students were invited to participate. The invitation asked them to be part of a study that examined their ability to diagnose their strengths and weaknesses with Spanish.

In an attempt to control for a homogeneous population of participants, only those students with the following criteria were included in the final data analysis: (1) students whose native language was English, (2) students who

completed the SA instrument, and for students in Advanced Spanish (3) students who completed the online diagnostic test.

3.3. Materials

3.3.1. Self-assessment questionnaire

The self-assessment factors for this study consisted of modified questions taken from the Reading, Writing, Listening, and Speaking Skills Self-Assessment Grid, which is a component of the DIALANG project. In the design of the SA instrument, specific course content was taken into consideration along with the direct experience learners had in practicing all four language skills (Ross, 1998). The questions for the present study were modified accordingly to reflect the program objectives for each level of instruction, and as Ross (1998) indicates, the “can-do” type items from the DIALANG may give more accurate results than other formats for SA. For the present study, the “can-do” statements are referred to as “skills-based” because they were taken from the criterion-referenced items of the DIALANG and then tailored to each skill across levels of instruction. The PI of this project obtained a course syllabus from corresponding course coordinators for each level and created a self-assessment questionnaire for each skill in each class. Following this, a meeting was held with each coordinator to discuss specific objectives for each skill within each class. Course coordinators were given the chance to independently review the questionnaire and were asked to be sure the items were tailored to specific course objectives for each skill. All course coordinators provided modifications to the questionnaires.

In the questionnaire, students were asked to indicate how they would rate their Spanish in each situation presented. They were told to circle the appropriate rating. The 5 point rating scale which ranged from “1” being “strongly disagree” to “5” being “strongly agree” was again tailored after the DIALANG scales. See Appendix A for sample items from the inventory used for self-assessment. The internal consistency (Cronbach’s α) for the SAs, across skill area and level of instruction, ranged from .77 to .92 ($M = .85$, $SD = .04$).

3.3.2. Ability instrument for advanced learners

The online diagnostic test was designed for learners of Advanced Spanish and consisted of a total of 100 questions. The test was divided into four different sections: vocabulary, grammar, reading, and listening. Students in Advanced Spanish were expected to achieve a total score of 75 or higher. The grammar section of the online diagnostic test consisted of 20 questions that varied in difficulty levels and included grammatical structures appropriate to each stage of instruction. The vocabulary section of the online diagnostic test consisted of 20 questions and reflected lexical items learned across beginning, intermediate, and advanced courses. The reading section of the online diagnostic test consisted of 30 questions. The reading section included eight different passages of varying styles and lengths with corresponding multiple choice questions. The content of the readings involved history, daily lives, personal anecdotes, and encyclopedia-like topics. All readings were in line with the ACTFL proficiency guidelines for advanced readers and consisted of familiar sentence patterns and topics. The listening section of the online diagnostic test consisted of 30 questions and included a variation of short dialogs and prose readings. Each listening item contained content that emphasized the daily lives of the students, historical events, or personal anecdotes.

All questions on the abilities test were in multiple choice formats and included 4 possible answers, and all incorrect answers were plausible. The test was an online test, and learners were given a time frame of 3 days to take the test from any computer at any location. Each section of the test was timed so that if students spent too much time on any question the items vanished. The scores for each section and the composite score was sent directly to the lead author of this study immediately after completion of the test.

4. Data analysis and results

Means, standard deviations and ranges were calculated for each skill and course level.

SA scores and online diagnostic test scores were correlated to determine the strength of their associations.

4.1. Research question one

When measured with a criterion-referenced instrument, what are the self-ratings of reading, writing, listening, and speaking skills with beginner, intermediate, and advanced learners of Spanish?

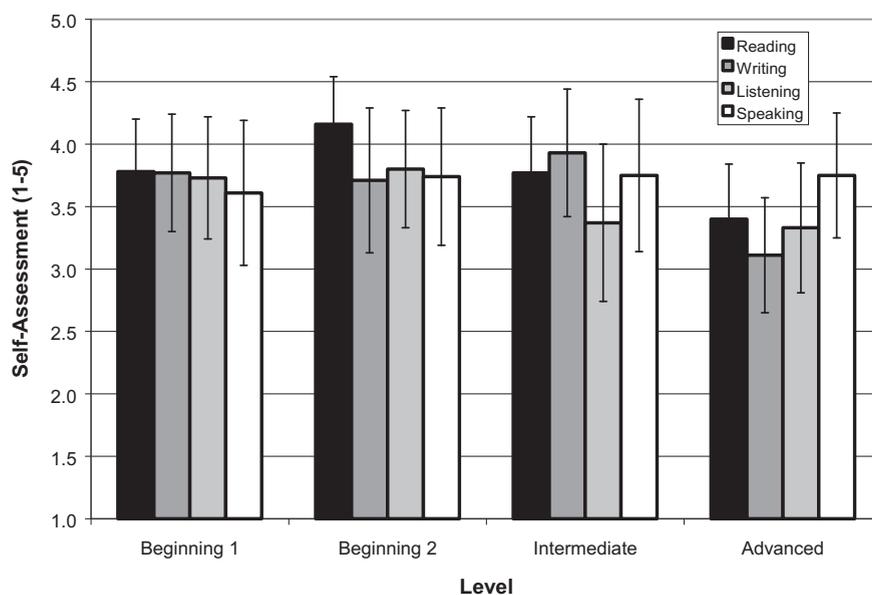


Fig. 1. Self-assessment ratings as a function of level and skill. The bars show means and the whiskers show standard deviations.

4.1.1. Beginning Spanish: Spanish 101 and 102

Beginning Spanish consists of both Spanish 101 (first semester) and 102 (second semester). Fig. 1 shows the means, standard deviations and ranges for each skill and course level; note that the numbers reflect item averages and not sums.

As shown in Fig. 1, learners in Spanish 101 ranked three abilities almost the same: read ($M = 3.78$), write ($M = 3.77$), and listen ($M = 3.73$), which indicates that they “agree” that they can use the language in each situation presented for each skill. They rated their ability to speak ($M = 3.61$) the lowest. The average score indicates a degree of neutrality with the different situations presented. The standard deviations for all four skills were quite low, which indicates that the data points tend to be close to the mean and that there is not much variation from the mean SA ratings within the group of beginners. This statement is further substantiated by the minimum and maximum values for each skill listed in Table 2.

Student scores ranged from “neutral” to “agree” with each situation. Further analysis of specific statements with high ratings is listed in Appendix B.

Learners in Spanish 102 ranked their ability to read the highest ($M = 4.16$) as did the students from Spanish 101. The students in Spanish 102 assessed their other abilities in the following order: listening ($M = 3.80$), speaking ($M = 3.74$), and finally writing ($M = 3.71$). Again, the standard deviations were quite low for all four skills, which indicate little variation within this group of learners. Listening had the highest range of scores (see Table 2), but a closer look at frequency data shows that only a few students had a mean rating of “disagree.” Further analysis of specific statements with high ratings can be found in Appendix B.

Table 2
Ranges for self-assessment by skill and level.

	Beginning 1			Beginning 2			Intermediate			Advanced		
	Min	Max	Range	Min	Max	Range	Min	Max	Range	Min	Max	Range
SA reading	3.04	4.46	1.43	3.68	4.62	.95	2.61	4.66	2.05	2.44	4.43	1.99
SA writing	2.90	4.36	1.46	2.70	4.46	1.76	2.5	4.76	2.28	2.32	4.47	2.15
SA listening	2.93	4.54	1.61	2.38	4.68	2.30	2.33	4.21	1.88	1.77	4.30	2.56
SA speaking	2.71	4.46	1.75	3.00	4.54	1.54	3.40	4.34	.95	1.58	4.2	.62

$n =$ Beginning 1 – 28; Beginning 2 – 37; Intermediate – 38; Advanced – 150.

4.1.2. Intermediate Spanish

As indicated in Fig. 1, learners in Spanish 201 ranked their abilities in the following order: writing ($M = 3.93$), reading ($M = 3.77$), speaking ($M = 3.75$), and listening ($M = 3.37$). Once again the low standard deviations reveal little variation from the mean for this group, although Table 2 shows the difference between the largest and smallest scores in the distribution for reading, writing, and listening to be 2 data points. A closer glance at frequency data shows, again, that only a few students are at the high and low rankings for these three abilities. A closer examination of itemized statements for each skill is listed in Appendix B.

4.1.3. Advanced Spanish

As indicated in Fig. 1, students in advanced Spanish rated their ability to speak the highest ($M = 3.75$), which is completely opposite of the beginner groups that rated their speaking abilities the lowest ($M = 3.61$), but it is important to note that each inventory was tailored to the specific objectives of each class. The students in advanced Spanish ranked their other abilities in the following order: reading ($M = 3.40$), listening ($M = 3.33$), and finally writing ($M = 3.11$). Standard deviations were low for each ability with advanced learners, indicating little variation with this group. The ranges for each ability are listed in Table 2. The range for speaking is very low, demonstrating a small difference in distribution of ratings for this ability. The range for speaking with advanced learners was overall the lowest range value for all groups and all abilities. Overall, learners' self-ratings of their ability over the three different instructional levels indicated that students tend to be "neutral" or "agree" with the statements about each situation presented in each item. Further analysis of specific statements with high ratings can be found in Appendix B.

4.2. Research question two

With advanced learners of Spanish, is there a relationship between criterion-referenced self-assessed ratings of skills and subsequent performance?

For subsequent performance, means and standard deviations were calculated for vocabulary, grammar, reading, and listening. Table 3 lists findings. The average scores for learners from the advanced level of language instruction achieved appropriate scores for this stage.

Table 4 lists the correlations among mean scores for SA ratings and scores on the online exam.

As we see in Table 4, the SA rating for listening, speaking, reading, and writing are positively related to scores on all components of the diagnostic test (vocabulary, grammar, listening, and reading). There was no oral component to the online diagnostic exam, but nonetheless, positive correlations were found with different scores on the diagnostic test and SA ratings for speaking abilities. This may underscore the fact that the different skills all depend on each other at the advanced levels, and that speaking ability may depend on the other abilities as well.

Furthermore, the total score for SA ratings positively associated with the total score on the diagnostic exam. Thus, it can be said that with advanced learners there is a positive relationship between criterion-referenced self-assessed ratings of all four skills and subsequent performance. Learners at the advanced stages of acquisition make self-assessments that are significantly related to their language abilities, although the magnitude of the relations is not substantial suggesting there may be important moderators to be identified. Findings will be discussed in the following section.

Table 3

Means and standard deviations for online diagnostic test with advanced learners.

	Advanced learners	
	Mean	SD
Vocabulary	16.07	2.83
Grammar	15.09	2.52
Listening	22.79	4.36
Reading	24.28	4.76
Total	78.05	

$n = 150$.

Table 4
Correlations — SA and online test with advanced learners.

	SA reading	SA writing	SA listening	SA speaking	Vocab test	Grammar test	Listening test	Reading test	Test total	SA total
SA reading	1	.594**	.697**	.522**	.329**	.207*	.226**	.229**	.298**	.842**
SA writing		1	.530**	.544**	.257**	.259**	.222**	.245**	.296**	.773**
SA listening			1	.695**	.290**	.262**	.274**	.209*	.311**	.871**
SA speaking				1	.252**	.306**	.251**	.211**	.288**	.768**
Vocab test					1	.584**	.542**	.757**	.830**	.332**
Grammar test						1	.442**	.576**	.699**	.305**
Listening test							1	.692**	.847**	.274**
Reading test								1	.912**	.239**
Online test total									1	.341**
SA test total										1

**Correlation is significant at the 0.01 level (2-tailed); *correlation is significant at the 0.05 level (2-tailed); $n = 150$.

5. Discussion

Findings for the first semester beginners, second semester beginners and the intermediate group appeared to be consistent with course goals for these levels of instruction as well as prior research (see results section for details by item). For example, the first-semester beginner group rated their reading, writing, and listening almost the same at “very good” ($M = 3.7$) with the lowest self-rating for speaking ($M = 3.6$). This is consistent with prior research (Ross, 1998) where speaking was rated the lowest with beginner language learners. For obvious reasons, students who are true beginners of a language tend to rate the silent skills higher as they are not producing oral output that can be judged and evaluated immediately by professors and peers. These self-ratings changed for second semester beginners as the reading score rose slightly to a mean of 4.16, with speaking increasing to an average rating of 3.7. Although this is not a substantial increase, it is worth noting that students rate their speaking higher as second semester beginners. These results echo De Saint-Leger (2009) who contends that student perceptions of their speaking abilities improve over time. Also, reading is rated highest for both beginning groups and this is expected given the extensive exposure learners have to written input via technology. In many language learning contexts reading predates experiences for listening and speaking (Ross, 1998).

A closer look at specific items on the SA instruments indicates that students in beginner Spanish agree that when they read they can understand the general idea of simple informational texts and short simple descriptions. They also confirmed that they understand better when they reread texts. These findings parallel the objectives on the course syllabus for Spanish 101 as students are taught to strategically read informational type texts that include a lot of supporting descriptions and details. Students are asked to read passages several times at home before coming to class the next day. Students also indicated a very good ability to fill in forms with personal details and to follow oral speech which is very slow and carefully articulated. Again, these tasks are emphasized in the class, so these ratings underscore the fact that the course may be meeting specific goals with particular abilities. Students also appear to be quite confident speaking about personal details such as name, nationality, major, etc, which again reinforces the fact that the course is meeting specific objectives. Students in the true beginner group rated their ability to read and understand texts with job-related language the lowest ($M = 3.0$, neutral), along with their ability to orally present an argument for their point of view ($M = 2.7$, neutral). These neutral ratings are expected given that these tasks are initiated toward the end of the semester in preparation for the upcoming course. With second semester beginners, the mean ratings for these two items increases to the next rating of “agree” ($M = 4.0$). Again, these ratings with beginners, as well as those with second semester beginners and intermediate learners, are in agreement with the goals of the courses, and they also may indicate that learners are gaining great awareness of their own strengths and weaknesses.

Future research should validate the SA instrument with actual performance for each skill with beginner and intermediate learners. The present study attempted to do this, but the 100 point diagnostic test did not include sufficient numbers of items tailored to beginner learners for accurate statistical analysis. The next step with this research would be to use this SA instrument with a well-developed and validated test for beginner and intermediate learners. We hypothesize that adult beginner learners of Spanish in the USA may under-evaluate themselves as they are masters of their first language, and therefore at the beginning stages they may underestimate their second language abilities. We need more research to validate this hypothesis. Future inquiry should also include a longitudinal study of

students who are true beginners to see if abilities to self-assess improve over time. More research is needed that examines the place and value of self-assessment for learner outcomes across levels of instruction that may not be part of the language major.

With advanced learners of Spanish, the present study validates the relationship between individual self-assessment of reading, writing, listening and speaking abilities and scores on an online abilities test.³ The associations for SA ratings for all four skills and scores for each component of the test are approximately the same strength. Prior research with ESL learners has shown that proficiency level does matter with accuracy of self-assessment, with more proficient students underestimating their abilities (Heilenman, 1990; Oskarsson, 1984). However, the present investigation uses correlations because the scales for the two forms of assessment are different, and therefore claims about overestimation or underestimation cannot be made. The strength of the association between SA and online test scores is not influenced by the absolute levels of the ratings or online scores. In the present study, however, overall self-assessment scores by ability positively associated with the overall online test scores, and it can be asserted that advanced students do in fact know when they are relatively better or poorer at different skills. The present study utilized a criterion-referenced instrument, and the wording on the questions, the type of questions, and also the second language may affect the outcomes.

In the present study, data-driven generalizations with the advanced learners are compatible with previous assertions that self-assessment instruments yield higher correlations with measures of proficiency if the SA items are specific and focused (Pierce et al., 1993). Brantmeier (2006) reported that learners at the advanced stage do not accurately assess reading with self-rated scales, whereas Brantmeier and Vanderplank (2008) also found that with criterion-referenced items learners accurately estimate reading abilities. Prior research, where researchers did not use criterion-referenced SA items, has also shown that this is not the case for less advanced learners where no significant results were found with relationships between self-assessment and tests (Blanche and Merino, 1989; Janssen-van Dieten, 1989; Pierce et al., 1993). SA test method effect could make the difference. Again, future research should utilize criterion-referenced SA items with less advanced learners to examine the relationships with abilities tests. Based on present results, it can also be postulated that the kind of SA instrument does matter with advanced learners.

The present study moves beyond prior research with advanced learners as it utilized a skills-based instrument and also includes all four skills: reading, writing, listening and speaking. A closer examination of the positive associations between overall SA ratings of each ability and different subskills (i.e. vocabulary and grammar) underscores the complexity of factors involved in the acquisition of different abilities. For example, the SA reading scores held positive associations with online reading, vocabulary, grammar and writing scores, and the SA writing scores held the same powerful associations. The SA listening rankings yielded positive associations with vocabulary, grammar and speaking, and the SA speaking ranking yielded the same significant relationships.⁴ Earlier findings with ESL participants indicate that learners more accurately assess speaking than writing (Rasch, 1979) and that ESL students can provide adequate assessments of their speaking and writing, but not their reading (Krausert, 1991). In the present study, the fact that the relative sizes of the correlations are all of about the same magnitude merits discussion in light or prior findings. The present study may demonstrate that the different skills all depend on each other at the advanced levels, and that the different skills are correlated facets of a single higher order construct-second language acquisition.

The development of SA instruments across stages of acquisition is quite labor-intensive and time-consuming at the onset because of the necessity to tailor each item to the course objectives. The present study shows that the ratings on SA criterion-references instruments do match specific course goals by ability. Once these SA tools are established, they may be very attractive to large language programs as they can be used each semester. The present study used class time for students to complete the SA inventory, but when used for assessment purposes, the tools could be completed online outside of class. Results could be made available to program directors for immediate analysis and reporting. An online SA tool may be an efficient and effective way to document the progress of individuals and groups of learners.

The implementation of the SA instrument was met with great enthusiasm from many individual students. Without soliciting feedback on the instrument, students emailed instructors stating that they really enjoyed being able to rate themselves as it gave them a stronger sense of responsibility for their own learning. This comment makes great sense as students are traditionally accustomed to filling out end of semester course evaluation where they judge the course,

³ The online abilities test did not include a speaking component; however, the present study does report findings of the associations between SA of speaking and other abilities.

⁴ It is not the intent of the present study to see which individual SA items hold the highest correlations with scores on different abilities; however, a future study could examine specific items within each SA category.

the instructor, and the syllabus. Students are not often asked to evaluate themselves and their own performance, and it appears that students welcome this type of responsibility. Several also mentioned that they would like to do this throughout the semester in addition to the end. This request is consistent with suggestions by Little (2009), who proposes that with appropriate SA instruments, students become equipped with skills to self-assess for each skill and consequently develop metacognitive skills and greater learner autonomy. Alderson (2005) also comments on how learners who participate in self-assessment activities may be able to develop more realistic goals for further language learning, and it may consequently contribute to students having more control over their language learning experiences.

From an administrative standpoint, utilizing an online SA instrument to examine learner outcomes was also received positively. The online SA test only takes approximately 20 min to complete, in contrast to the 90–100 min usually needed for a sit-down abilities test. Also, there is no need to set up a testing schedule with rooms, and instructors do not have to grade or report grades. Many instructors commented on how beneficial the SA instrument is for assessment purposes as it offers further corresponding details about specific course objectives with detailed feedback. The SA instrument also involves little anxiety on the part of the learner and instructor, and the lack of affective barriers was specifically noted by program coordinators and directors. These assertions are consistent with Cohen (1984) where students report issues involved in taking language tests. Overall, the SA instrument is an important component of the overall assessment portfolio.

Future research should also consider self-assessment and other background factors and sociolinguistic variables, such as native language and ethnicity. Many learners in the USA are multilingual, and learners from different cultures may self-assess in different ways. For example, Matsumo (2009) reported that Japanese EFL students may assess themselves more harshly than might be expected because of cultural issues with modesty. Likewise, Luoma (2004) points out that variations with SA finding may be regional in origin. The present study omitted all learners whose first language was not English in order to account for homogeneity of participants. A future study should examine scores by first language. However, as suggested by Alderson (2005), the size of the sample may be small and consequently results should be interpreted with caution. The present study used English as the testing language for SA items, and this may also have an impact on findings with the participants who were excluded from final data analysis.

6. Conclusion

The present study shows that, with criterion-referenced SA instruments that are tailored to course objectives, language programs may be able to document how well individual students learn and develop over time. Students may also benefit from involvement in the assessment process. Additionally, the present study demonstrates that advanced students know when they are relatively better or poorer at different skills. Criterion-referenced SA has proven to be a powerful low-stakes assessment tool that may benefit both the individual learners as well as the language program. Finally, SA can provide a valuable departure from traditional testing formats that helps the learner become aware of individual strengths and weaknesses.

Acknowledgments

A version of this paper was presented at AAAL in Chicago, 2011. A special thanks to the audience for very helpful suggestions and comments, and in particular, thanks to Dr. Andrew Cohen for his insightful feedback during the discussion at AAAL. Thanks to the anonymous reviewers who provided very useful feedback.

Appendix A. Sample items from self-assessment instrument

Beginning Spanish (101)

Reading

1. I can understand the general idea of simple informational texts and short simple descriptions, especially if they contain pictures which help to explain the text.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can understand very short, simple texts, putting together familiar names, words and basic phrases, by for example rereading parts of the text.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can understand very simple sentences, for example on notices and posters or in catalogs.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can understand a simple personal letter in which the writer tells or asks me about aspects of everyday life.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can understand everyday signs and notices in public places, such as streets, restaurants, railway stations and in workplaces.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Writing

1. I can write a short simple postcard, for example sending holiday greetings.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can fill in forms with personal details, for example writing my name, nationality and address on a hotel registration form.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can write short, simple notes and messages about everyday matters and everyday needs.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can write a very simple personal letter, for example thanking someone for something.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can write simple texts on topics which are familiar or of personal interest.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Listening

1. I can understand the most common words and very simple phrases about myself, people I know and things around me when people speak slowly and clearly.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can follow speech which is very slow and carefully articulated, with long pauses for me to get the meaning.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can understand enough to manage simple, routine conversations without too much effort.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can understand questions and instructions and follow short, simple directions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can understand the main points of clear standard speech on familiar matters connected with work, school, leisure, etc.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Speaking

1. I can speak about personal details, for example my name, nationality, major, etc.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I participate in simple conversations about everyday matters and everyday needs.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can talk about topics which are familiar or of personal interest.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can talk about and describe personal experiences and impressions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can talk about a wide range of subjects related to my interests.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

*Advanced Spanish (SPN 307)**Reading*

1. I can understand a personal letter in which the writer describes events, feelings and wishes.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can understand in detail a wide range of long, complex texts provided I can reread difficult sections.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I understand a broad range of vocabulary, but I sometimes experience difficulty with less common words or phrases.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can read short stories, following the flow of thoughts and actions and thus understanding the overall meaning and many details.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can understand in a narrative the motives for the characters' actions and their consequences for the development of the plot.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Writing

1. When I write, I can speculate about causes, consequences and hypothetical situations.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can write clear and well-structured texts and express my points of view at some length, and presenting relevant supporting detail.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. When I write, I can give clear detailed descriptions of complex subjects.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. When I write, I can synthesize information and arguments from a number of sources.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can usually write without consulting a dictionary.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Listening

1. I can understand longer stretches of speech and lectures and follow complex lines of argument, provided the topic is reasonably familiar and the speech is clear and in standard language.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. In TV and current-affairs programs or programs of personal or professional interest, I can understand the main points provided the speech is relatively slow and clear and/or the visual material supports the commentary.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can understand simple directions relating to how to get from X to Y, by foot or public transit.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can guess the meaning of occasional unknown words from the context and understand sentence meaning if the topic discussed is familiar.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can follow clear speech in everyday conversation, though in a real life situation I will sometimes have to ask for repetition of particular words and phrases.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Speaking

1. I can speak about topics that are reasonably familiar by using standard language.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I can discuss TV and current-affairs programs or programs of personal or professional interest.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I can ask for simple directions relating to how to get from X to Y, by foot or public transit.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I can participate in everyday conversation, though in some situations I am unable to produce particular words and phrases.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I can talk about causes, consequences and hypothetical situations.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Appendix B. Statements with highest ratings by skill and level of instruction

4.1.1. Beginning Spanish: Spanish 101

Reading	“I can understand the general idea of simple informational texts and short simple descriptions, especially if they contain pictures which help to explain the text”	[M = 4.4]
Reading	“I can understand very short, simple texts, putting together familiar names, words and basic phrases, by for example rereading parts of the text”	[M = 4.4]
Reading	“I can understand texts that contain everyday or job-related language”	[M = 3.0]
Reading	“I can search one long or several short texts to locate specific information I need to help me complete a task”	[M = 3.0]
Writing	“I can fill in forms with personal details, for example writing my name, nationality and address on a hotel registration form”	[M = 4.3]
Writing	“I can write letters highlighting the personal significance of events and experiences”	[M = 2.9]
Writing	“I can construct a basic chain of reasoned argument”	[M = 2.9]
Listening	“I can follow speech which is very slow and carefully articulated, with long pauses for me to get the meaning”	[M = 4.5]
Listening	“In TV and current-affairs programs or programs of personal or professional interest, I can understand the main points provided the speech is relatively slow and clear and/or where the visual material supports the commentary”	[M = 2.9]
Speaking	“I can speak about personal details, for example my name, nationality, major, etc.”	[M = 4.5]
Speaking	“I can report, pass on information, and present some arguments for or against a particular point of view.”	[M = 2.7]

4.1.1. Beginning Spanish: Spanish 102

Reading		[M = 4.16]
Reading	“I can understand the general idea of simple informational texts and short simple descriptions, especially if they contain pictures which help to explain the text”	[M = 4.6]
Reading	“I can search one long or several short texts to locate specific information I need to help me complete a task”	[M = 3.7]
Listening		[M = 3.80]
Listening	“I can understand the most common words and very simple phrases about myself, people I know and things around me when people speak slowly and clearly”	[M = 4.6]
Listening	“I can follow speech which is very slow and carefully articulated, with long pauses for me to get the meaning”	[M = 4.6]
Listening	“I can understand standard spoken language, live or broadcast, on a variety of topics. Only extreme background noise, unclear structure and/or idiomatic usage causes some problems”	[M = 2.4]
Speaking		[M = 3.74]
Speaking	“I can speak about personal details, for example my name, nationality, major, etc.”	[M = 4.5]
Speaking	“When I speak, I can speculate about causes and consequences”	[M = 3.0]
Writing		[M = 3.71]
Writing	“I can fill in forms with personal details, for example writing my name, nationality and address on a hotel registration form”	[M = 4.5]
Writing	“I can write an essay or report, passing on information and presenting some arguments for or against a particular point of view”	[M = 2.7]

4.1.2. Intermediate Spanish: Spanish 201

Writing		[M = 3.93]
Writing	“I can write simple texts on topics which are familiar or of personal interest”	[M = 4.7]
Writing	“I can usually write without consulting a dictionary”	[M = 2.5]
Reading		[M = 3.77]
Reading	“I can understand very simple sentences, for example on notices and posters or in catalogues”	[M = 4.6]
Reading	“I can understand a simple personal letter in which the writer tells or asks me about aspects of everyday life, or in which the writer describes events, feelings and wishes”	[M = 4.6]
Reading	“I can go quickly through long and complex texts, locating relevant details”	[M = 2.6]
Speaking		[M = 3.75]
Speaking	“I can speak about topics which are familiar or of personal interest.”	[M = 4.3]
Speaking	“When I speak, I can synthesize information and arguments”	[M = 3.4]
Listening		[M = 3.37]
Listening	“I can guess the meaning of occasional unknown words from the context and understand sentence meaning if the topic discussed is familiar”	[M = 4.2]
Listening	“I can keep up with an animated conversation between native speakers”	[M = 2.3]
Listening	“I can recognize a wide range of idiomatic expressions and colloquialisms and recognize changes in style”	[M = 2.3]

4.1.3. Advanced Spanish

Speaking		[M = 3.75]
Speaking	“I can ask for simple directions relating to how to get from X to Y, by foot or public transit”	[M = 4.5]
Speaking	“I can talk about a variety of abstract and concrete topics with simple sentence structure”	[M = 4.5]
Speaking	“When I speak, I can use a wide range of idiomatic expressions and colloquialisms”	[M = 2.3]
Reading		[M = 3.40]
Reading	“I can understand a personal letter in which the writer describes events, feelings and wishes”	[M = 4.4]
Reading	“I can understand specialized language in articles and technical instructions even if they are not in my field”	[M = 2.4]
Listening		[M = 3.33]
Listening	“I can understand simple directions relating to how to get from X to Y, by foot or public transit”	[M = 4.3]
Listening	“I have no difficulty in understanding any kind of spoken language, whether live or broadcast, delivered at fast native speed”	[M = 1.8]
Writing		[M = 3.11]
Writing	“When I write, I can speculate about causes, consequences and hypothetical situations”	[M = 4.2]
Writing	“I can write so well that my texts cannot be improved significantly even by teachers of writing”	[M = 1.6]

References

- Alderson, C.A., 2005. *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment*. Continuum, New York, NY.
- Bachman, L.F., Palmer, A., 1981. The construct validity of the FSI oral proficiency interview. *Language Learning* 31, 67–86.
- Bachman, L.F., Palmer, A., 1982. The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16, 449–465.
- Bachman, L.F., Palmer, A., 1996. *Language Testing in Practice*. Oxford University Press, Oxford.
- Baniabdelrahman, A.A., 2010. The effect of the use of self-assessment on EFL students' performance in reading comprehension in English. *TESL-EJ* 14, 1–22.
- Birckbichler, D., Corl, K., Deville, C., 1993. The dynamics of language program testing: implications for articulation and program revision. In: *The Dynamics of Language Program Direction*. Heinle & Heinle, Boston, MA.
- Blanche, P., Merino, B., 1989. Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning* 39, 313–340.
- Brantmeier, C., 2005. Nonlinguistic variables in advanced L2 reading: learner's self assessment and enjoyment. *Foreign Language Annals* 38, 493–503.
- Brantmeier, C., 2006. Advanced L2 learners and reading placement: self-assessment, computer based testing, and subsequent performance. *System* 34 (1), 15–35.
- Brantmeier, C., Vanderplank, R., 2008. Descriptive and criterion-referenced self assessment with L2 readers. *System* 36 (3), 456–477.
- Buck, G., 1992. Listening comprehension: construct validity and trait characteristics. *Language Learning* 42, 313–357.

- Butler, Y., Lee, J., 2006. On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal* 90, 506–518.
- Cohen, A.D., 1984. On taking language tests: what the students report. *Language Testing*, 70–81.
- De Saint-Leger, D., 2009. Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals* 42, 158–178.
- De Saint-Leger, D., Storch, N., 2009. Learners' perceptions and attitudes: implications for willingness to communicate in an L2 classroom. *System* 37 (2), 269–285.
- Deville, M., Deville, C., 1999. Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273–299.
- DIALANG project. Retrieved from: <http://www.lancs.ac.uk/researchcenterprise/dialang/about>.
- Ferguson, N., 1989. Self-assessment of listening comprehension. *International Review of Applied Linguistics* 126, 149–156.
- Hargan, N., 1994. Learner autonomy by remote control. *System* 22 (4), 455–462.
- Heilenman, L.K., 1990. Self-assessment of second language ability: the role of response effects. *Language Testing* 7, 174–201.
- Janssen-van Dielen, A., 1989. The development of a test of Dutch as a second language: the validity of self-assessments by inexperienced subjects. *Language Testing* 6, 30–46.
- Krausert, S.R., 1991. Determining the usefulness of self-assessment of foreign language skills: post-secondary ESL students' placement contribution. PhD. dissertation, University of Southern California.
- LeBlanc, R., Painchaud, G., 1985. Self-assessment as a second language placement instrument. *TESOL Quarterly* 19, 673–687.
- Little, D., 2009. Language learner autonomy and the European language portfolio: two L2 English examples. *Language Teaching: Surveys and Studies* 42, 222–233.
- Luoma, S., 2004. Self-assessment in DIALANG: an account of test development. In: Milanovic, M., Weir, C. (Eds.), *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference, July 2001*, pp. 143–157.
- Matsumo, S., 2009. Self-, peer- and teacher-assessments in Japanese University EFL writing classrooms. *Language Testing* 26 (1), 75–100.
- Oskarsson, M., 1978. *Approaches to Self-assessment in Foreign Language Learning*. Council of Europe, Council for Cultural Cooperation, Strasbourg, France.
- Oskarsson, M., 1984. *Self-assessment of Foreign Language Skills: A Survey of Research and Development Work*. Council for Cultural Cooperation, Strasbourg, France.
- Pierce, B.M., Swain, M., Hart, D., 1993. Self-assessment, French immersion, and locus of control. *Applied Linguistics* 14, 25–42.
- Rasch, G., 1979. *Probabilistic Models for Some Intelligence and Achievement Tests*. Danish Institute for Educational Research, Copenhagen (Expanded edition, 1980. University of Chicago Press, Chicago).
- Ross, S., 1998. Self-assessment in second language testing: a meta-analysis of experimental factors. *Language Testing* 15, 1–20.
- Stern, H.H., 1983. *Fundamental Concepts of Language Teaching*. Oxford.
- Wan-a-rom, U., 2010. Self-assessment of word knowledge with graded readers: a preliminary study. *Reading in a Foreign Language* 22, 323–338.
- Wongsotorn, A., 1981. Self-assessment in English skills by undergraduate and graduate students in Thai universities. In: Read, J. (Ed.), *Directions in Language Testing*. Regional Language Centre, Singapore.