



Descriptive and criterion-referenced self-assessment with L2 readers

Cindy Brantmeier^{a,*}, Robert Vanderplank^b

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

^b *Oxford University Language Centre, Oxford University, 12 Woodstock Road, Oxford OX2 6HT, UK*

Received 20 August 2007; received in revised form 7 March 2008; accepted 14 March 2008

Abstract

Brantmeier [Brantmeier, C., 2006. Advanced L2 learners and reading placement: self-assessment, computer-based testing, and subsequent performance. *System* 34 (1), 15–35] found that self-assessment (SA) of second language (L2) reading ability is not an accurate predictor for computer-based testing or subsequent classroom performance. With 359 advanced L2 university-level learners, the present study attempts to provide new empirical evidence concerning the use of SA, when measured via criterion-referenced items, as a predictor of reading scores on a computer-based placement test and subsequent reading achievement. Before reading, participants completed a contextualized, criterion-referenced instrument for SA that was created from the European Language Portfolio. A four-item contextualized SA questionnaire was also completed after reading. Participants completed the following: self-assessment questionnaire taken prior to computer-based test, computer-based test of 8 readings, 2 in-class readings, written recall, multiple-choice, topic familiarity questionnaire, and post-reading SA questionnaire. Findings indicated that a descriptive and criterion-referenced questionnaire for SA may be a reliable predictor of both reading scores on a computer-based test and subsequent classroom performance when comprehension is measured via sentence completion and multiple choice items. The criterion-referenced questionnaire did not prove to be a reliable predictor of reading comprehension with the written recall task for the short story. The post-reading SA instrument proved reliable for all three comprehension tasks. Results hold important implications for both language placement directors and test constructors.

© 2008 Elsevier Ltd. All rights reserved.

* Corresponding author.

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

Keywords: Second language reading; Self-assessment; Criterion-referenced; Advanced language learners

1. Introduction

Traditionally, L2 reading was an ability that many language instructors took for granted, in part because it was viewed as a silent, passive skill. Since the early 1980s research in reading has revealed that it is a complex process involving multiple, interacting factors. In fact, the most comprehensive L2 reading model to date illuminates the importance of both linguistic and affective variables (Bernhardt, 2005), however, the role of self-assessment in L2 reading with adults has yet to be fully explored. A clearer understanding of the role of self-assessment in the L2 reading process would help practitioners with language placement and curriculum development, and it will also aid teachers in raising students' awareness of their own reading strengths and weaknesses.

Recently, Brantmeier (2006) contended that placement directors cannot depend on SA of reading ability, as measured via non-descriptive self-assessment items, with advanced learners who are entering the university. In that study, the self-assessment items asked the readers to rate themselves as a reader of Spanish on a five point scale from “not good at all” to “excellent”. The SA findings are not surprising given that language learners in the USA receive more feedback on their productive abilities (writing and speaking) than they do on their reading and listening abilities. Some researchers advocate using SA instruments instead of standardized exams for language placement in university programs (Krausert, 1991), and others advocate using SA as a valuable tool to aid in placement decisions (LeBlanc and Painchaud (1985). Given the findings of Brantmeier (2006), along with prior research on self-assessment and placement, the need for a descriptive and criterion-referenced SA instrument for use with advanced placement in reading is needed. If SA items are going to be used to assist in placement decisions, learners need to know, in uncomplicated and straightforward terms, exactly what it is they are trying to assess. Currently, there exists a need to expand research on placement testing with advanced learners that incorporates aspects beyond language components, such as self-assessment.

2. Literature review

Research on self-assessment across disciplines has dealt with issues of measurement, emphasizing reliability and construct validity (Ross, 1998). Investigations concerning self-assessment in language learning in general have examined the value of SA in proficiency testing with participants of all ages (see Blance and Merino, 1989; Ross, 1998 for detailed reviews), and to date the majority of research has utilized subjects at the initial stages of English studies (Bachman and Palmer, 1981, 1982, 1996; Buck, 1992; Ferguson, 1989; LeBlanc and Painchaud, 1985; Wongsotorn, 1981). Results with these learners show that SA positively correlates with language abilities. Oscarson (1997) offers a thorough review of investigations concerning SA in the second and foreign language context and highlights the dearth of research in this area (Falchikov and Boud, 1989; Heidt, 1979; and Oscarson, 1984). With learners of advanced Spanish, Brantmeier

ier (2005) examined self-assessment of L2 reading abilities and general enjoyment of L2 reading. Before completing reading, students were asked to rate themselves as readers of Spanish, and possible answers included five possible choices from 1 (I think that I am an excellent reader of Spanish) to 5 (I do not think that I am a good reader of Spanish at all).

Results indicated that both enjoyment levels and self-assessment correlated with comprehension when measured via a written recall task, but not when assessed via a multiple-choice task. The author suggested that future research of this nature should include measures of self-assessment *after* reading performance as well as an additional reading comprehension task (besides written recall and multiple-choice items). Additionally, it was suggested that future research might examine the relationship between self-assessment, reading placement test achievement, and subsequent reading performance. Brantmeier (2006) utilized the same self-assessment inventory before participants completed an online reading test for language placement and an in-class reading test. Additionally, the author utilized a questionnaire of self-assessment ratings for use *after* students completed the reading in class in order to contextualize items and gain an assessment of comprehension of the specific reading utilized for the study (Schraw et al., 1995; Tobias and Everson, 1998; Everson, Chief Research Scientist, College Board of USA, personal communication, November, 2004). Furthermore, as part of the in-class reading test, participants completed three different tasks to measure comprehension. Results revealed no positive associations between SA, as measured before and after reading via a 5-point scale, and comprehension, as measured via an online placement test and an in-class reading test.

Table 1 highlights select studies on self-assessment and language learning. As depicted, results are conflicting across stages of acquisition. The value of self-assessment as a factor in language learning and language placement at the university is an issue that merits further inquiry.

Table 1
Selected prior studies on SA and language learning abilities

Author	Participants	Results
Oscarson (1978)	Adult students of EFL	SA positively correlated with teacher ratings and written test scores
LeBlanc and Painchaud (1985)	First-year students of English and French as L2	Positive correlations between SA and standardized proficiency tests for listening and reading
Krausert (1991)	Adult students of ESL	Positive results for SA of reading, writing and speaking abilities
Hargan (1994)	Adult students of ESL	A multiple-choice test and SA test indicated same level of placement
Birckbichler et al. (1993)	Adult students of ESL	SA correlated higher than any other variable with comprehension scores
Ross (1998)	Adult students of L2	Meta-analysis revealed the largest number of correlations between SA and L2 reading
Deville and Deville (1999)	Adults of L2 in USA	Self-assessment is an effective starting point for CAT
Brantmeier (2005)	Adults of Spanish as L2	SA positively correlated with enjoyment and written recall, but not with multiple-choice items
Brantmeier (2006)	Adults of Spanish as L2	Results indicated that SA is not an accurate predictor of reading on a computer-based test and subsequent reading

Alderson (2005) details how SA is a central component of DIALANG, a language diagnostic/placement test for 15 different European languages (see www.dialang.org). The DIALANG view (p. 264) is that learners should be able to compare their SA ratings with their performance in any particular skill in low-stakes testing environments so that the discrepancies revealed may provide valuable insights into their learning and beliefs about language learning. Findings of the pilot tests in reading indicated that while there was a significant relationship between self-assessed level of reading and levels of reading in terms of items linked to the Common European Framework (CEFR), there were also marked differences according to background variables such as mother tongue, age, sex, length of time studying English and frequency of use. In pilot English tests, higher level learners were better able to self-assess than lower level learners (p. 137). Crucially, DIALANG teams found that sub-skills in reading such as the ability to make inferences were not tied to different CEFR levels. Given that inferencing can refer to skills from guessing to working out meaning from contextual and co-textual clues, this finding may not be altogether surprising, especially if texts are designed to elicit inferencing.

The reliability of SA has also been considered indirectly in studies of language learners' attributions of success and failure. Graham (2004), investigating the performance of 500 teenage students in French, found discrepancies between learners' self-assessed lack of confidence in French and the marks they had obtained in school tests. Graham suggests that they were assessing themselves on how well they thought they could perform in French rather than on the good marks they obtained. Covington (2000, p.174) offers a plausible explanation for this in terms of performance goals vs. learning goals: 'learning goals refer to increasing one's competency, understanding, and appreciation for what is being learned. . . performance goals. . . involve outperforming others as a means to aggrandize one's ability status at the expense of peers'.

3. The present study

The Department of Romance Languages and Literatures at a Midwestern private university utilizes an online placement test to assess freshmen during the summer months prior to arrival at the university. The computer-based test immediately reports scores upon completion to both the exam taker and program director and the score is used for decisions regarding placement. While the computer-based test has proved reliable for most placement cases, there are always some cases with marginal scores where the learner scores on the "margin", or directly in between two levels. The placement director must then decide if the learner should enroll in the lower or higher level. Directors would find another means of assessment useful in helping to determine the placement of these learners. Because of time constraints at the beginning of the semester with students who need to enroll in language classes before all sections are full, the director must act quickly on the placement of students with marginal scores. Reliable self-assessment would provide an efficient means of placing students with marginal reading scores.

Spanish is the fastest growing major at the university. The majority of freshmen place into advanced Spanish, which marks the beginning of the Spanish major or minor. Furthermore, many students take Spanish courses without declaring the major or minor areas of study. The Spanish major emphasizes the reading of lengthy, authentic texts written by

authors of the Spanish speaking world. Obviously, given this focus, reading is an important skill.

The present study differs from Brantmeier (2006) in the following ways: (1) present study uses a more robust population of participants, (2) present study utilizes criterion-referenced self-assessment items in addition to descriptive items, (3) present study uses two different reading passages instead of one for in-class reading, (4) present study assesses comprehension via three tasks instead of two. The following research questions guide the present study:

1. Do pre-test SA ratings of reading, as measured with both a descriptive question and criterion-referenced instrument, accurately predict achievement on a computer-based test?
2. Do pre-test SA ratings of reading, as measured with a criterion-referenced instrument, accurately predict subsequent reading performance?
3. Do post-reading SA ratings of L2 reading comprehension accurately reflect achievement?
4. Does reading performance on a computer-based test for placement predict subsequent reading performance?

3.1. Participants

Participants in the present study were 359 students, approximately 118 men and 241 women, ages 19–22, enrolled in an advanced level Spanish grammar and composition course at a middle-sized, private university in the United States. Many freshmen place directly into this advanced level course, which marks the first course toward a major in Spanish and is the first in a two-course sequence taken immediately before entering the literature courses. According to ACTFL proficiency guidelines, students at this level should be able to read longer prose of several paragraphs in length, and comprehension depends not only on situational and subject matter knowledge but on increasing control of the language. During the course students are assigned to read authentic literary works from the literary canon. Because there is no language requirement at the university, all students enrolled in the course voluntarily. In other words, students are not obliged to take language courses and anyone can choose to study languages.

For the present study, data were collected during four different semesters. 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 who achieved the appropriate composite score on the computer-based test (tested into Advanced Spanish), (2) students whose native language was English, (3) students who enrolled in Advanced Spanish the semester immediately after taking the computer-based test, and (4) students who completed all tasks for both data collection settings.

3.2. Self-assessment questionnaire

The self-assessment factors for this study consisted of modified questions taken from several different sources: (1) the Reading Skills Self-Assessment Grid, which is

a component of the European Language Portfolio developed under the direction of the European Council; (2) Oscarson's (1984) self-assessment questionnaire, and; (3) LeBlanc and Painchaud (1985) statements for SA. The questions for the present study were modified accordingly to reflect the program objectives for reading and consisted of 16 items. For the present study, the term "self-assessment" refers to the judgments or beliefs that learners make about their L2 learning abilities and performance, including assessment for placement (Oscarson, 1989). SA is a format that requires students to rate their own language abilities or competence, whether through performance-ability self-assessments, comprehension self-assessments, or observation self-assessments (Brown and Hudson, 2002, p. 83). See Appendix A for the complete inventory used for self-assessment.

A self-assessment question asked students which description most closely matched their comprehension level when reading a text in Spanish. This 6 point rating scale ranged from "I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues" to "I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works". Following this question, the participants indicated on a Likert scale from 1 to 5 whether they strongly disagreed or strongly agreed with 15 contextualized and criterion-referenced statements. Questions included items such as "I can understand a very short narrative on everyday topics if written in simple language", and "I can read a variety of factual and literary texts and am able to understand the content (main topics and the different points of view)". The readers completed the pre-test self-assessment questions online before beginning a timed placement examination that determined their level of Spanish, and consequently this may be considered a high-stakes assessment situation. For the DIALANG project, Alderson specifically excludes SA from high stakes testing. He regards SA more as a valuable descriptive and explanatory tool for providing feedback to learners. Brown and Hudson (2002) echo these concerns with criterion-referenced testing. The purpose of the present study is to examine whether SA criterion-referenced items could potentially be a valuable additional component to an otherwise high-stakes test for learners who score on the margins. For the current investigation, criterion-referenced testing means "... any test that is primarily designed to describe the performances of examinees in terms of the amount that they know of a specific domain of knowledge or set of objectives" (Brown and Hudson, 2002, p. 5).

As mentioned earlier, Brantmeier (2006) utilized a questionnaire of self-assessment ratings for use *after* students completed the reading in class, in order to contextualize items and gain an assessment of comprehension of the specific readings. No positive associations between SA ratings and comprehension were found. Consequently, the post-test self-assessment questions utilized for the present study were modified. The first question asked the reader how well he understood the passage he had just read. The five possible answers ranged from "Not very well at all" to "Very well". The following questions asked how well the participants did on each specific assessment task, for example: "How well do you think you did on the written recall task for the passage you just read?" Again, the five possible answers ranged from "Not very well at all" to "Very well".

3.3. Reading section for computer-based test

There are eight different readings on the computer-based test which includes vignettes of varying styles and lengths. Topics and styles include the following: daily lives of students, historical vignettes, a poem, personal narratives, and encyclopedia-like readings. Comprehension is tested via multiple-choice items with four possible answers: one correct response and three distractors. All distractors were plausible (Alderson, 2000; Bernhardt, 1991; Wolf, 1993), and some questions included inferential items. All multiple-choice questions were written in Spanish.

3.4. Reading passages for classroom performance

Unlike Brantmeier (2006), which utilized one lengthy short story in a study to measure subsequent classroom performance, the present study included two different readings during class: a narrative and an essay. Both readings were on paper, not on the computer. The narrative was a 512 word excerpt taken from a story by Rigoberta Menchu, which takes place in Guatemala and details the life of a *quiche* woman. The essay contained approximately 318 words and is about television, culture and diversity. More specifically, it explored different perspectives of the term “culture”.

In the present study, steps were taken to control for gender differences in topic familiarity (Bügel and Buunk, 1996; Brantmeier, 2002, 2003, 2006; Chavez, 2001; Schueller, 1999). Before the investigation, the reading passages for the present study were piloted with 103 students to ensure that they contained a topic familiar to students. All students read the passage and then indicated levels of topic familiarity. Results revealed no significant gender differences in topic familiarity, with both male and female participants indicating that “I was familiar with some and unfamiliar with some” of the passage topic.

3.5. Assessment tasks for classroom performance

Research has not yet revealed the perfect test to measure reading comprehension (Alderson, 2000), and to be able to generalize, a variety of assessment tasks are needed to measure comprehension (Bernhardt, 1991). For the present study, three different assessment tasks were utilized: written recall, sentence completion, and multiple-choice. All three tasks were completed in English, the native language of participants.

According to Bernhardt (1991), the written recall assessment task offers the purest depiction of comprehension, as tester interference and retrieval cues do not exist. For the present study, immediately after reading each passage, the written recall protocol followed standard procedures and asked readers, without looking back at the passage, to recall and write down as much as they could of what they had just read (Bernhardt, 1983; Brantmeier, 2002; Carrell, 1983; Lee, 1986a; Young and Oxford, 1997; and more).

Researchers utilize a variety of scoring rubrics for the recall task that include idea units, propositional units, pausal units, etc. (Barnett, 1988; Brantmeier, 2002; Carrell, 1983; Lee, 1986a,b; among others), and most agree that the pausal unit is the most effective. A pausal unit is a unit or entity that during normally paced oral reading has a pause on each end of it (Bernhardt, 1991). To analyze the texts and recalls, the present study utilizes the pausal

unit protocol. Oral reading by four native speakers was used for the initial division of pausal units for each text, and then written recall protocols were checked for the presence of absence of such units. The total number of correct pausal units was utilized for the written recall score.

Both sentence completion and multiple-choice tasks place limits on answers. Sentence completion is an open-response task where all possible answers are foreseeable. The objectivity of scoring depends upon the comprehensiveness of the answer key. Multiple-choice questions are the most popular means of assessing reading comprehension. All answers are pre-determined and answers are either right or wrong. For each of the multiple choice questions four possible responses were created (Alderson, 2000; Bernhardt, 1991; Wolf, 1993). In addition, another provision was also included: test-takers were not able to determine correct responses by looking at the other questions on the page. Each sentence-completion item solicited the same information as a multiple-choice question, and both global and local questions were included.

To avert a test of writing instead of reading, the written recall was completed in the reader's native language (Alderson, 2000; Bernhardt, 1991; Lee and Ballman, 1987; Wolf, 1993). Additionally, all assessment tasks for subsequent reading performance were completed in the reader's native language.

3.6. *Topic familiarity*

A topic-familiarity questionnaire was completed by all participants. Topic familiarity was assessed via questions with five possible choices ranged from 1 (I was really familiar with this topic) to 5 (I was not familiar with this topic at all).

3.7. *Data collection procedures*

First, all students were asked to assess their L2 reading abilities, and then they completed the computer-based test for placement. In order to determine subsequent reading performance, the same students participated in an investigation during regular class time during the 3rd week of class. During a regular class period of 50 min, all subjects completed the following instruments in this order for each passage: reading passage, written recall, sentence-completion items, multiple-choice questions, post self-assessment questionnaire, and topic-familiarity questionnaire.

It is important to note that when participants were invited to participate in an experiment concerning L2 reading, they were only told that they would complete the above tasks. No other details about the experiment were provided to participants. No participants declined to participate in the study. The researcher and/or a research assistant along with all instructors for the courses were present during data collection sessions to ensure that students did not look back at any previous pages while reading and completing all tasks.

3.8. *Data analysis*

The two different self-assessment ratings were the independent variables: the descriptive SA and criterion-referenced SA. The dependent variable for the reading section of the

computer-based test was the total number of correct responses. There were three different dependent variables for subsequent reading performance: recall, sentence completion and multiple choice. For recall, the researcher and an additional rater identified the total pausal units for the text. The interrater reliability index was .95. The total number of pausal units for the short story was 77, and for the essay the total number was 68. For sentence-completion questions, several possible correct answers were formulated for each question. Separately the raters identified all possible correct responses, and the interrater reliability index was .96. For multiple-choice questions, there was one correct answer for each question.

For each research question, means, standard deviations, ranges, and median scores were calculated. To assess the degree to which SA scores predicted reading scores on the computer-based test and subsequent reading performance, data were examined using regression analysis. The regression procedure determined the strength of these relationships as well as the amount of variance explained by self-assessment (Brantmeier, 2006). The Alpha level for statistical significance was set at .05. In addition, data were analyzed to see which SA item has the highest bivariate correlation with comprehension when only significant SA predictors are entered into the regression equation. A regression model was calculated to determine which combination of significant predictors correlated with comprehension.

4. Results

The mean score on the 6 point pre-self-assessment item (descriptive) was 3.8 and standard deviation .63. This indicates an average SA of “I can read articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints. I can understand contemporary literary prose”. It is important to note that the scale ranged from “I can understand familiar names, words and very simple sentences. . .” to “I can read with ease virtually all forms of the written language, including

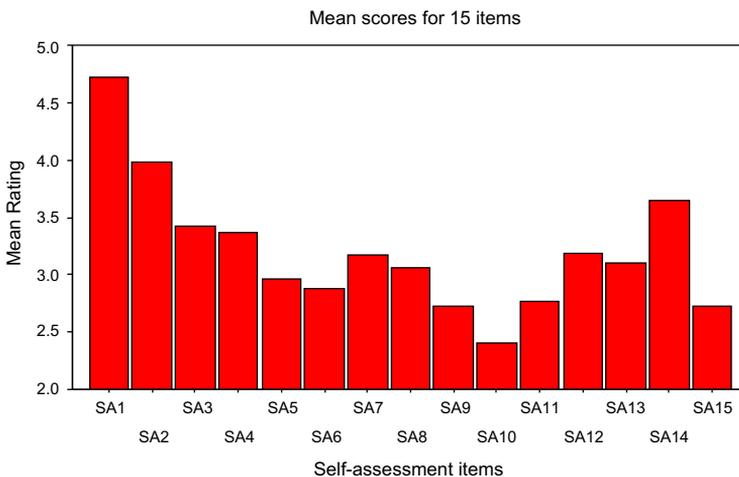


Fig. 1. Self-assessment of L2 reading ability: mean scores for 15 items.

Table 2

Mean scores and standard deviations for reading on computer-based test (CBT) and subsequent reading achievement

	CBT	Recall		SC		Multiple-choice	
		Txt1	Txt2	Txt1	Txt2	Txt1	Txt2
Possible score	30.0	77.0	68.0	10.0	10.0	10.0	10.0
Mean	24.2	14.4	11.6	5.0	7.4	7.9	6.8
SD	4.7	5.6	5.2	1.8	1.2	1.4	1.0
Minimum	4.0	5.0	1.0	1.0	1.0	3.0	1.0
Maximum	29.0	30.0	29.0	10.0	10.0	10.0	10.0
Range	25.0	25.0	28.0	9.0	9.0	7.0	9.0

$n = 359$; Txt 1 = short story; Txt 2 = essay.

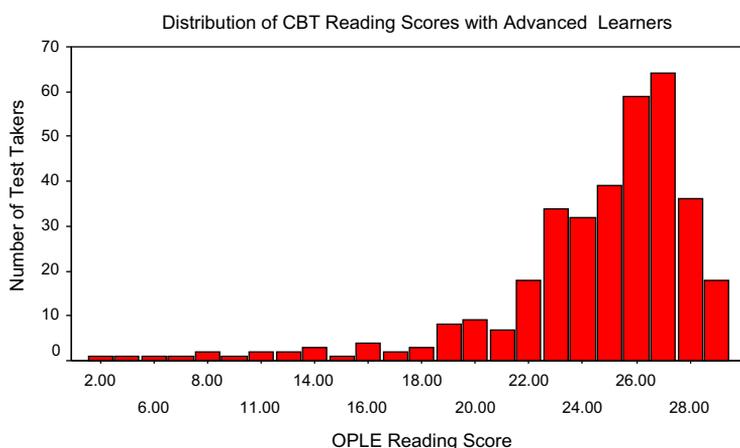


Fig. 2. Computer-based test for placement: reading scores.

abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works”. See Fig. 1 for mean scores for the 15 criterion-referenced SA items.

Again, “1” indicates that they strongly disagree and “5” indicates that they strongly agree with the statement. The question with the highest mean rating (strongly agree) was “I can understand a personal letter in which the writer tells or asks me about aspects of everyday life”. The question with the lowest mean rating (strongly disagree) was “I can understand texts written in a very colloquial style and containing many idiomatic expressions or slang”. Most of the ratings fell between 2.5 and 3.5 indicating a “neutral” response to the SA items.

Mean scores and standard deviations for the computer-based test and subsequent performance (three assessment tasks) are listed in Table 2. As listed on the table, participants scored a mean score of 24.2, or an average of 81% on the reading items for the computer-based test. Fig. 2 graphically illustrates the distribution of reading scores on the computer-based test with these advanced learners.

Table 3

Regression analysis: association between self-assessment and comprehension

Predictors (constant)	R^2	T -Ratio	P
<i>Reader's self-assessment: pre-reading</i>			
CBT			
SA (descriptive item)			
Multiple-choice	0.04	13.16	0.00*
SA (15 criterion items)			
Multiple-choice	0.11	7.24	0.00*
Short story			
SA (10 criterion items)			
Recall	0.02	2.21	0.60
Sent-completion	0.08	3.23	0.00*
Multiple-choice	0.10	12.84	0.00*
Essay			
SA (8 criterion items)			
Recall	0.02	4.56	0.67
Sent-completion	0.01	6.30	0.90
Multiple-choice	0.02	8.38	0.74
<i>Reader's self-assessment: post-reading</i>			
Short story			
SA (2 items)			
Recall	0.06	1.46	0.00*
Sent-completion	0.23	4.13	0.00*
Multiple-choice	0.05	17.82	0.00*
Essay			
SA (2 items)			
Recall	0.22	2.42	0.00*
Sent-completion	0.26	5.70	0.00*
Multiple-choice	0.03	19.33	0.00*

* $p < 0.05$; $n = 359$.

For subsequent performance on Text 1, participants scored a mean score of 14.4 on recall, or 19% of the total possible pausal units. With this same passage, participants achieved a score of 50% on sentence-completion and 80% on multiple-choice. With Text 2, participants scored a mean score of 11.6%, or 17% of the total possible pausal units. With this same passage, participants achieved a score of 70% on both sentence-completion and multiple-choice.

A descriptive SA score was regressed on total multiple-choice for eight different reading texts that were part of a computer-based test used for placement. Table 3 lists the associations between self-assessment and comprehension.

Findings indicated that the pre-test SA descriptive rating positively relates to total multiple-choice score on the reading from the computer-based test ($t = 13.16$, $p < .05$). Additionally, the 15 pre-test SA items were regressed on total multiple-choice for the eight different reading texts on the computer-based test. The 15 combined items positively related to total multiple-choice scores ($t = 7.24$, $p < .05$). In addition, data were further analyzed to see which SA item had the highest bivariate correlation with comprehension when only significant SA predictors were entered into the regression

Table 4
Regression analysis: association between SA items and comprehension

Predictors (constant)	R^2	T -Ratio	P
<i>Reader's self-assessment pre-reading</i>			
<i>CBT</i>			
Multiple-choice			
SA Item # 14	0.06	16.38	0.00*
SA Items # 14, 7	0.10	16.82	0.00*
<i>Short story</i>			
Sentence-completion			
SA Item # 14	0.04	6.99	0.00*
SA Item # 14, 2	0.06	3.10	0.00*
SA Item # 14, 2, 7	0.07	3.26	0.00*
Multiple-choice			
SA Item # 7	0.05	56.30	0.00*
SA Item # 7, 14	0.08	19.57	0.00*
<i>Reader's self-assessment post-reading</i>			
<i>Short story</i>			
Recall			
SA Item "well recall"	0.06	2.40	0.00*
Sentence-completion			
SA Item "well SC"	0.22	6.76	0.00*
Multiple-choice			
SA Item "well undrstd"	0.04	20.20	0.00*
SA Item "well undrstd + well MC"	0.05	17.82	0.00*
<i>Essay</i>			
Recall			
SA Item "well recall"	0.21	4.52	0.00*
SA Item "well recall + well undrstd"	0.22	2.41	0.00*
Sentence-completion			
SA Item "well SC"	0.24	10.70	0.00*
SA Item "well sentence-completion + well undrstd"	0.26	5.70	0.00*
Multiple-choice			
SA Item "well MC"	0.11	11.78	0.00*
SA Item "well MC + well undrstd"	0.12	8.39	0.00*

* $p < 0.05$; $n = 359$; SC = sentence-completion; MC = multiple-choice.

equation. Table 4 lists the positive associations between specific combinations of SA items and comprehension.

As indicated, SA Item # 14, "I can understand contemporary and classical literary texts of short stories" positively associated with the score on the computer-based test and had the highest correlation with comprehension. The combination of this item with #7 "I can go beyond the concrete plot of a narrative and grasp implicit meanings, ideas and connections" also yielded a positive association to score on the computer-based test.

In order to examine whether the criterion-referenced SA items accurately predicted subsequent reading performance, selected SA items were regressed on recall, sen-

tence-completion and multiple-choice for the two different reading texts (short story and essay). With Text 1 (short story), 10 different SA criterion items were regressed on all three tasks. The 10 SA criterion items were selected because of their direct relevance to reading a short story. Findings are listed in Table 3. Results yielded a positive association between self-assessment and sentence completion ($t = 3.23, p < .05$), and self-assessment and multiple choice ($t = 12.84, p < .05$). With Text 2 (essay), eight criterion items were regressed on all three tasks. The eight items were selected because they directly related to reading an essay. No positive associations were found. Data were further analyzed to see which SA item held the highest bivariate correlation with comprehension when only significant SA predictors are entered into the regression equation. With the short story, the item that positively related to recall was the same as the association for the computer-based test, SA Item #14 (“I can understand contemporary and classical literary texts of short stories”). Additionally, with the short story for sentence completion two different combinations of items positively related: (1) SA Item #14 and SA Item #2, “I can read a variety of factual and literary texts and am able to understand the content (main topics and the different points of view),” and (2) SA Item #14, SA Item #2, and SA Item # 7 “I can go beyond the concrete plot of a narrative and grasp implicit meanings, ideas and connections”. For the multiple choice items with Text 1 (short story), SA Item #7 provided the highest correlation. The only combination of SA items with a positive association with multiple choice was SA Item #7 and SA Item #14 (“I can understand contemporary and classical literary texts of short stories”).

A regression analysis was also calculated to determine the association between the post-test SA items and comprehension scores. Findings are listed in Table 3. Results show a positive association between post SA items and all three comprehension assessment task for both texts. Table 4 shows the item-specific associations. A scatterplot revealed that the relationship between post-self assessment ratings and subsequent performance on recall, SC, and multiple-choice for both texts was linear. Consequently, to see whether one variable is predictable from the other, a Pearson’s Correlation was calculated. Results for both texts are reported in Tables 5 and 6.

With both texts, findings indicated positive correlations between Post-test Item 1 “How well did you understand the reading passage in Spanish?” and Post-test Items 2 and 4 for how well they think they did on the specific comprehension assessment tasks.

Table 7 lists the relationships between computer-based test scores and subsequent reading performance. As indicated, the computer-based test score predicted performance on both the sentence-completion and multiple-choice items for the short story, and the computer-based test score predicted performance on recall, SC, and multiple-choice for the essay.

5. Discussion and implications

The following summarizes significant findings:

A significant association for the Descriptive Pre-reading SA and computer-based test reading score was found. For the Criterion-Referenced SA rating, a positive association was found for both the computer-based reading score and subsequent reading performance with Text 1 (story) (for multiple-choice and SC), but not for Text 2 (essay).

Table 5
Correlations Post SA and comprehension for short story

		RECALL	SC	MULTIPLE-CHOICE	WELL UNDRSTD	WELL RECALL	WELL SC	WELL MULTIPLE-CHOICE
RECALL	Pearson correlation Sig. (2-tailed) N	1 . 359	.303(**) .000 359	.033 .537 359	.162(**) .002 359	.243(**) .000 359	.180(**) .001 359	.660(**) .000 359
SC	Pearson correlation Sig. (2-tailed) N	.303(**) .000 359	1 . 359	.429(**) .000 359	.261(**) .000 359	.412(**) .000 359	.471(**) .000 359	.279(**) .000 359
MULTIPLE-CHOICE	Pearson correlation Sig. (2-tailed) N	.033 .537 359	.429(**) .000 359	1 . 359	.188(**) .000 359	.173(**) .001 359	.236(**) .000 359	.178(**) .001 359
WELL UNDRSTD	Pearson correlation Sig. (2-tailed) N	.162(**) .002 359	.261(**) .000 359	.188(**) .000 359	1 . 359	.545(**) .000 359	.428(**) .000 359	.347(**) .000 359
WELL RECALL	Pearson correlation Sig. (2-tailed) N	.243(**) .000 359	.412(**) .000 359	.173(**) .001 359	.545(**) .000 359	1 . 359	.533(**) .000 359	.305(**) .000 359
WELL SC	Pearson correlation Sig. (2-tailed) N	.180(**) .001 359	.471(**) .000 359	.236(**) .000 359	.428(**) .000 359	.533(**) .000 359	1 . 359	.405(**) .000 359
WELL MULTIPLE-CHOICE	Pearson correlation Sig. (2-tailed) N	.660(**) .000 359	.279(**) .000 359	.178(**) .001 359	.347(**) .000 359	.305(**) .000 359	.405(**) .000 359	1 . 359

** Correlation is significant at the 0.01 level (2-tailed).

Table 6
Correlations Post SA and comprehension for essay

		REC	SC	MULTIPLE- CHOICE	WELL UNDRSTD	WELL RECALL	WELL SC	WELL MULTIPLE- CHOICE
REC	Pearson correlation	1	.442(**)	.317(**)	.365(**)	.456(**)	.391(**)	.307(**)
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000
	N	358	358	357	357	357	356	356
SC	Pearson correlation	.442(**)	1	.306(**)	.372(**)	.385(**)	.494(**)	.336(**)
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000
	N	358	358	357	357	357	356	356
MULTIPLE- CHOICE	Pearson correlation	.317(**)	.306(**)	1	.291(**)	.239(**)	.317(**)	.329(**)
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000
	N	357	357	357	357	357	356	356
WELL UNDRSTD	Pearson correlation	.365(**)	.372(**)	.291(**)	1	.607(**)	.577(**)	.542(**)
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000
	N	357	357	357	357	357	356	356
WELL RECALL	Pearson correlation	.456(**)	.385(**)	.239(**)	.607(**)	1	.681(**)	.524(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000
	N	357	357	357	357	357	356	356
WELL SC	Pearson correlation	.391(**)	.494(**)	.317(**)	.577(**)	.681(**)	1	.626(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000
	N	356	356	356	356	356	356	356
WELL MULTIPLE- CHOICE	Pearson correlation	.307(**)	.336(**)	.329(**)	.542(**)	.524(**)	.626(**)	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.
	N	356	356	356	356	356	356	356

** Correlation is significant at the 0.01 level (2-tailed).

Table 7

Regression analysis: computer-based test for placement and subsequent performance

Predictors (constant)	R^2	T -Ratio	P
<i>CBT reading score</i>			
Short story recall	0.01	2.50	0.07
SC	0.05	5.32	0.00*
MC	0.06	14.72	0.00*
Essay recall	0.03	4.43	0.00*
SC	0.02	6.90	0.00*
MC	0.03	8.47	0.00*

$p < 0.05$; $n = 359$; SC = sentence-completion; MC = multiple-choice.

Additionally, the Post-reading SA yielded a positive association with all three comprehension tasks completed for the essay and the short story. Pre-test criterion-referenced SA item #14 proved to be the most reliable (“I can understand contemporary and classical literary texts”). In other words, students who strongly agreed or agreed with this statement scored highest on written recall, sentence-completion, and multiple-choice items. As indicated, some SA items lose predictive power when new variables are entered into the equation. In the end, SA item #14 shows the strongest predictive power in terms of the computer-based test for placement. The same item holds greatest effects on the Short Story for both the sentence-completion and multiple-choice. It is striking to note that students score high on multiple-choice and sentence-completion but very low on recall. Recall is a production task that relies heavily on memory. Recall is known to be the ‘pure’ form of testing reading comprehension, but it appears that it does not test something which taps the sort of recognition and understanding which sentence completion and multiple-choice do. This may also suggest why recall did not correlate well with the short story. While the SA ‘I can understand contemporary ...’ is directly related to the multiple-choice and sentence completion, there is actually no SA item ‘I can recall short stories and texts after reading them’. A future study of this nature may want to include an SA item that directly assesses recall.

In a closer examination of the test scores and the SA scores for different components, there seems to be a discrepancy between the test scores and what might be called the ‘confidence rating’ of students’ SA on the measures of SA used. The median score for the computer-based test was 25 with a large cluster of 55% gaining 25–29 (max. score 30). As the graph (Fig. 2) shows, there are no other comparable clusters. The students appear to be doing better on the computer-based test than their largely neutral self-ratings would indicate. In other words, in terms of test scores, high-performing students are tending to underestimate their performance. This finding is consistent with results of Kruger and Dunning (1999), who found that high-performing students tend to underestimate their performance, while low-performing students often over-estimate their competence. However, this finding conflicts with those of the DIALANG project where learners at higher levels were better able to self assess than learners from lower levels. Research on the role of attribution in language learning (for example, Graham, 2004) would suggest that a possible reason for this lies in students being able to make a distinction between what they perceive their ‘real life’ performance in a foreign language to be and what they achieve in tests. Even learners at an advanced level may feel that they cannot approach certain authentic Spanish texts with the same confident

assumptions of understanding that truly proficient users of the language are able to. An over-confident response to ‘I can understand contemporary and classical literary texts’, might be rewarded with a text of truly Joycean complexity in Spanish in the subsequent class.

In the present study, as in Brantmeier (2006), the reliability of the reading scores on a computer-based test in predicting subsequent reading performance was also examined. As discussed earlier, in Brantmeier’s (2006) experiment, 71 participants completed the online placement exam and read only one text during class to measure subsequent performance (a vignette from a short story). The low scores on comprehension tasks were similar to the present study, and the computer-based test again predicted actual reading performance in the course. Unfortunately, the weakness of any computer-based test for placement that examines different skills is that it may place weak readers into the higher levels of instruction because the test takers may score very high on listening and grammar, and therefore the differences balance out. The total composite score determines course placement. Overall, the present findings partially echo earlier results in that readers testing into the higher levels of instruction may not have the skills to cope with advanced level texts. Additionally, findings in the present study are similar to Brantmeier (2006) with great variation of reading abilities within the participants (see Fig. 2). It is important to note that the present study included two different reading passages instead of one for subsequent performance, but the readings on the computer-based test were exactly the same. When looking closer at the DIALANG project, it appears that more factual texts are used which may mean less engagement between reader and text. This difference may be more important in terms of self-assessment.

Brantmeier (2006) used non-descriptive items for the pre-test self assessment rating and found no significant associations between SA and reading scores for the computer-based test and between SA and subsequent performance. The present study used a more descriptive, criterion-referenced self-assessment instrument with the same level of readers. With the descriptive pre-reading SA item (see Appendix A), results revealed a positive association with reading performance (with the computer-based test and classroom investigation). This finding holds important implications for the design of a pre-test SA inventory to be used in placement with advanced learners. This descriptive question could be used in addition to reading scores on the computer-based test to place advanced learners. Additionally, positive associations were found between the pre-test criterion-referenced SA items and reading scores with the short story, but not with the essay. Given that these results appear dependent on the text style (short story vs. essay), generalizations cannot be made at this time. However, if a language program emphasizes the reading of narratives and short stories, the finding may hold important implications. A program of this kind could use criterion-referenced SA item “I can understand contemporary and classical literary texts” as this item proved significant.

While Brantmeier (2006) found no significant correlations between post-test SA ratings and comprehension scores, the present study, which utilized different, more contextualized post-test SA items, yielded significant associations between post-test SA ratings and subsequent performance. This finding is crucial because the post-test SA instrument may be a reliable option for placing marginal readers. It must be noted that the students in the present study had the benefit of three weeks of classes before they

completed the post-test SA instrument. Perhaps their awareness of their own strengths and weaknesses in reading comprehension has been raised during this period. A future study could test the value of post-test SA after the computer-based test used for placement. The positive findings discussed above, along with these results, could be taken into account when developing a SA inventory for placement. For example, the pre-test descriptive SA item, the pre-test criterion-referenced SA item, along with the post-test SA items could be the instrument used to aid placement decisions. This complete instrument would need to be tested in a high-stakes situation of placement to determine reliability. At this point, there is not enough evidence to validate the use of this instrument for placement.

As these are learners at an advanced level, we would expect a relatively high association between SA and test scores on most measures (Alderson, 2005). Taking this point further, the findings show the complexity of the issues involved in both SA and testing reading as different tests tend to reveal different aspects of the reading skill (different sub-skills). In this respect, then, we should expect some level of divergence. The present study revealed conflicting results for self-assessment with the different comprehension tasks (recall, sentence completion, and multiple-choice). Test-method effect for comprehension with self-assessment merits further inquiry. Also, as Alderson indicates (2005, 128), reading sub-skills, such as inferencing, are not directly related to the level of ability or textual difficulty.

Finally, the findings of the present study suggest that the nature of self-assessment has useful but ultimately limited reliability for reading placement. This may be due to the lack of exposure students in the USA have to self-assessment tasks, or as Kruger and Dunning (1999) have pointed out, their inherent unreliability even after training. Students are not often asked to evaluate or self-monitor progress or abilities in the language learning classroom. SA may be a very useful tool for raising awareness on the part of students, but it may not be a reliable predictor for placement. Again, the “high-stakes” issue of placement may also be an issue with SA. Both Alderson (2005) and Brown and Hudson (2002) advocate the use of SA in *non-testing* situations. However, given the fact that in the past self-assessment has been used for placement issues in the USA (Deville and Deville, 1999) and given the need for additional factors to determine language placement of marginal students (beyond a computer-based test), a need for the present study was substantiated.

Above all, findings of the present study call for additional research in this area. Future research could compare criterion-referenced SA results to background variables such as gender, and years of language study. Furthermore, many international graduate students in the USA are required to take an English language placement exam prior to enrolling in courses of study. A future study should examine criterion-referenced self-assessment by native language and proficiency with ESL learners. Finally, researchers could examine the role of self-assessment in metacognitive abilities of language learners. A large component of metacognition is the ability to self-assess, or self-evaluate. As shown in the conflicting results of previous work on self-assessment, learner ability to self-assess language skills is not commonplace or automatic. Consistent feedback on metacognitive strategies may help instructors obtain more reliable self-assessment results. Future inquiry could delve into this complex matter and offer corresponding instructional suggestions.

6. Conclusion

The present study revealed that pre-test SA can be a reliable criterion for reading placement when it is measured with criterion-referenced items, and certain SA items could be used in combination with the computer-based test for placement of advanced learners. Furthermore, these SA criteria could aid in the placement of readers with marginal reading scores (readers who score on the border between intermediate and advanced). Moreover, SA can help instructors diagnose strengths and weaknesses of individual readers. Assessing reading ability accurately may also lead to greater self-awareness and could contribute to goals to produce lifelong L2 readers. Readers should be taught to reflect critically on their L2 reading abilities on a regular basis, and this self-evaluation should be linked with the skills they are working on. Finally, the present investigation shows the value of SA in predicting performance on language placement tests and performance in class. As indicated, further research is needed to address the validity and reliability of self-assessment in the testing, or high-stakes, process.

Appendix A. SA items

A.1. Pre self-assessment item (descriptive)

Which of the following six descriptions most closely matches your skill level when you *read Spanish*? If you feel your skills levels are somewhere between two or more descriptions, check both descriptions. Please circle the most appropriate number.

1. I can understand familiar names, words and very simple sentences, for example, on notices and posters or in catalogues.
2. I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables, and I can understand short, simple personal letters.
3. I can understand texts that consist mainly of high-frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
4. I can read articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints. I can understand contemporary literary prose.
5. I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialized articles, even when they do not relate to my field.
6. I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works.

A.2. Pre self-assessment item (criterion referenced)

Please indicate your level of agreement with the following statements. Circle your choice.

1. 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

2. I can read a variety of factual and literary texts and am able to understand the content (main topics and the different points of view).

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

3. I can read and understand articles and reports on current problems in which the writers express specific attitudes and points of view.

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

4. I can read longer articles on current topics, distinguishing between facts, opinions and conclusions.

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

5. I can read novels, plays and 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

6. I can understand in a narrative or play 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

7. I can go beyond the concrete plot of a narrative and grasp implicit meanings, ideas and connections.

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

8. I can understand fairly long demanding texts and summarize them.

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

9. I can read contemporary literary texts with ease.

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

10. I can recognize the social, political or historical background of literary work.

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

11. I can recognize plays on words and appreciate texts whose real meaning is not explicit (for example irony, satire).

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

12. I can understand texts written in a very colloquial style and containing many idiomatic expressions or slang.

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

13. I can read texts such as literary columns or satirical glosses where much is said in an indirect and ambiguous way and which contain hidden value judgments.

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

14. I can understand contemporary and classical literary texts.

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

15. I can recognize different stylistic means (puns, metaphors, symbols, connotations, ambiguity) and appreciate and evaluate their function within the text.

Post Self-assessment Items 1. How well did you understand the reading passage in Spanish?

1	2	3	4	5
Not very well at all.	Not very well.	Ok	Well	Very Well

2. How well do you think you did on the written recall task for the Spanish passage you just read?

3. How well do you think you did on the sentence completion task for the Spanish passage you just read?

4. How well do you think you did on the multiple choice task for the Spanish passage you just read

References

- Alderson, C.A., 2000. *Assessing Reading*. Cambridge University Press, United Kingdom.
- 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.
- Barnett, M.A., 1988. Reading through context: how real and perceived strategy use affects L2 comprehension. *Modern Language Journal* 72, 150–160.
- Bernhardt, E.B., 1983. Syntactic and lexical/semantic skill in foreign language reading comprehension: the immediate recall protocol. *Die Unterrichtspraxis* 16, 27–33.
- Bernhardt, E.B., 1991. *Reading Development in a Second Language*. Ablex Publishing Corporation, Norwood, NJ.
- Bernhardt, E.B., 2005. Progress and procrastination in second language reading. *Annual Review of Applied Linguistics* 25, 133–150.
- 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.
- Blance, P., Merino, B., 1989. Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning* 39, 313–340.
- Brantmeier, C., 2002. The effects of passage content on second language reading comprehension by gender across instruction levels. In: Hammadou Sullivan, J. (Ed.), *Research in Second Language Learning*. Information Age Publishing, Greenwich, CT.
- Brantmeier, C., 2003. Does gender make a difference? Passage content and comprehension in second language reading. *Reading in a Foreign Language* 1 (15), 1–24.
- Brantmeier, C., 2005a. Nonlinguistic variables in advanced L2 reading: learner's self-assessment and enjoyment. *Foreign Language Annals* 38 (4), 493–503.
- Brantmeier, C., 2005b. Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension. *The Modern Language Journal* 89 (1), 37–53.
- Brantmeier, C., 2006. Advanced L2 learners and reading placement: self-assessment, computer-based testing, and subsequent performance. *System* 34 (1), 15–35.
- Brown, J.D., Hudson, T., 2002. *Criterion-referenced Language Testing*. Cambridge University Press, Cambridge, UK.
- Buck, G., 1992. Listening comprehension: construct validity and trait characteristics. *Language Learning* 42, 313–357.
- Bügel, K., Buunk, B.P., 1996. Sex differences in foreign language text comprehension: the role of interests and prior knowledge. *Modern Language Journal* 80, 15–31.
- Carrell, P.L., 1983. Some issues in studying the role of schemata, or background knowledge, in second language comprehension. *Reading in a Foreign Language* 1 (2), 81–92.
- Chavez, M., 2001. *Gender in the Language Classroom*. Multiple-choice Graw Hill, New York.
- Council of Europe, 2005. *Language Policy Division European, Language Portfolio (ELP)* <<http://culture2.coe.int/portfolio/inc.asp?L=E&>>.
- Covington, M.V., 2000. Goal theory, motivation, and school achievement: an integrative review. *Annual Review of Psychology* 51, 171–200.
- Deville, M., Deville, C., 1999. Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273–299.
- European Language Portfolio Self-Assessment <http://www-user.uni-bremen.de/~jsuther/self_assessment_reading.html>.
- Falchikov, N., Boud, D.J., 1989. Student self-assessment in higher education: a meta-analysis. *Review of Educational Research* 59 (4), 395–430.
- Ferguson, N., 1989. Self-assessment of listening comprehension. *International Review of Applied Linguistics* 126, 149–156.

- Graham, S., 2004. Giving up on modern foreign languages? Students perceptions of learning French. *The Modern Language Journal* 88 (2), 171–191.
- Hargan, N., 1994. Learner autonomy by remote control. *System* 22 (4), 455–462.
- Heidt, E., 1979. Self-evaluation in learning: a report on trends, experiences, and research finding. UNESCO, Division of Structures, Content, Methods, and Techniques of Education, Paris.
- Krausert, S.R., 1991. Determining the usefulness of self-assessment of foreign language skills: post-secondary ESL students' placement contribution. PhD. Diss., University of Southern California.
- Kruger, J., Dunning, D., 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77, 1121–1134.
- LeBlanc, R., Painchaud, G., 1985. Self-assessment as a second language placement instrument. *TESOL Quarterly* 19 (4), 673–687.
- Lee, J.F., 1986a. On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition* 8, 201–212.
- Lee, J.F., 1986b. Background knowledge and L2 reading. *Modern Language Journal* 70, 350–354.
- Lee, J.F., Ballman, T.L., 1987. Learners' ability to recall and rate important ideas of an expository text. In: VanPatten, B., Dvorck, T.R., Lee, J.F. (Eds.), *Foreign Language Learning: a Research Perspective*. Newbury House, Rowley, MA, pp. 108–117.
- Oscarson, M., 1978. Approaches to Self-assessment in Foreign Language Learning. Council of Europe, Council for Cultural Cooperation, Strasbourg.
- Oscarson, M., 1984. Self-assessment of foreign language skills: a survey of research and development work. Council of Europe, Council for Cultural Cooperation, Strasbourg.
- Oscarson, M., 1989. Self-assessment of language proficiency: rationales and applications. *Language Testing* 6, 1–13.
- Oscarson, M., 1997. Self-assessment of foreign and second language proficiency. In: *The Encyclopedia of Language and Education*, vol. 7. Kluwer Academic Publishers, pp. 175–187.
- Ross, S., 1998. Self-assessment in second language testing: a meta-analysis of experimental factors. *Language Testing* 15, 1–20.
- Schraw, G., Bruning, R., Svoboda, C., 1995. Sources of situational interest. *Journal of Reading Behavior* 27 (1), 1–17.
- Schueller, J., 1999. The effect of two types of strategy training on foreign language reading comprehension. An analysis by gender and proficiency. Unpublished doctoral dissertation. The University of Wisconsin, Madison.
- Tobias, S., Everson, H.T., 1998, April. Research on the assessment of metacognitive knowledge monitoring. Paper Presented at a Symposium on "Metacognition: Assessment and Training", at the annual convention of the American Educational Research Association, San Diego, CA.
- Wolf, D., 1993. A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal* 77, 473–489.
- 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.
- Young, D.J., Oxford, R., 1997. A gender-related analysis of strategies used to process input in the native language and in a foreign language. *Applied Language Learning* 8, 43–73.