



Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance

Cindy Brantmeier *

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

Received 10 February 2005; received in revised form 23 June 2005; accepted 11 August 2005

Abstract

There is a dearth of research involving advanced second language (L2) learners in the USA [Brantmeier, C., 2001. Second language reading research on passage content and gender: challenges for the intermediate level curriculum, *Foreign Language Annals* 34, 325–333; Young, D.J., 2003. Workshop on overcoming anxiety in the foreign language classroom. *Language Learning Series*, Washington University in St. Louis]. At some universities in the USA many freshmen test directly into the advanced levels of Spanish language instruction where the goal is to prepare students for the reading of lengthy, authentic texts. With 71 advanced L2 learners of Spanish, the current research project attempts to determine the reliability of self-assessment, as determined by a questionnaire previously utilized for placement [Birckbichler, D., Corl, K., Deville, C., 1993. The dynamics of language program testing: implications for articulation and program revision. *The Dynamics of Language Program Direction*. Heinle and Heinle, Boston, MA; Deville, M., Deville, C., 1999. Computer-adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273–299], to predict reading performance via computer-based testing (CBT) and subsequent reading achievement. All incoming freshmen completed self-assessment questions about L2 reading abilities before completing an online placement exam (OPLE). In order to assess subsequent reading performance, all students who tested into Advanced Spanish Grammar and Composition participated in an investigation during regular class time during the 3rd week of class. Participants read a short story, completed three different comprehension assessment tasks and additional self-assessment items. Overall, results of the present study suggest that self-assessment of L2 reading ability, as measured before and after reading via a 5-point scale, is not an accurate predictor variable for placement or subsequent performance. Advanced learners did not accurately estimate their L2 reading abilities

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

E-mail address: cbrantme@artsci.wustl.edu.

as measured via CBT and in-class performance. Findings do not offer conclusive evidence about the value of self-assessment as a predictor of performance on CBT or as an indicator of subsequent classroom performance, but the current investigation provides indications as to the direction in which research on self-assessment of L2 reading needs to move. A more contextualized, criterion-referenced self-assessment instrument may be more beneficial for the placement of advanced readers in the USA. These results both echo and contradict prior research, which calls attention to the need for more investigations concerning instrumentation of self-assessment as a factor to be used in placement decisions for advanced learners.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Self-assessment; Second language reading; Computer-based testing; Advanced second language learners; Language placement

1. Introduction

Placement testing is an essential part of university level language departments across the USA, and there is a current need to expand research on placement testing with advanced learners that incorporates aspects beyond linguistic components, such as self-assessment. For some time now, researchers have conducted investigations that examine issues concerning language placement tests (Bachman, 1990; Bachman and Palmer, 1996; Bernhardt et al., 2004; Byrnes, 1991; Chapelle, 2001; Dunkel, 1991; Heilenman, 1991; Lange et al., 1992; Schwartz, 1985), but to date it appears that none have specifically examined the relationship of the following variables with advanced language learners: self-assessment of reading abilities, CBT reading achievement, and subsequent reading performance. Given the emphasis on reading at the advanced levels of language instruction in the USA, an investigation of this nature is valuable for program directors that are responsible for placement testing and making decisions based on results.

Placement procedures vary greatly and include multiple variables that range from paper-and-pencil tests, online tests, oral interviews, and combinations of different procedures (Wheritt and Clearly, 1990). Besides test results, other variables often involved in placement decisions in the USA are years of language study prior to the university, time spent in target language countries, and scores on national standardized exams such as the advanced placement exam (AP) administered by the College Board. An issue of great importance regarding placement testing is that of practicality, and therefore many universities develop their own CBT for placement that reflects the objectives of their language programs. Even though online placement tests take a considerable amount of time and money to create, the practicality often justifies the costs. Conventional placement tests require testing schedules, appropriate types of rooms, proctors, graders, interviews, and more. Once created and validated, CBTs offer advantages that conventional placement tests do not. Chalhoub-Deville (2001) offers an overall review of what has transpired in the development of CBT in educational contexts and cites examples of institutions that are utilizing CBT for different purposes. Recently, Bernhardt et al. (2004) examined the practicality and efficiency of the CBT designed for language placement at Stanford University and reported that placement testing via the internet can be reliable and can match the objectives of the foreign language curriculum (p. 364), and in that study they specifically emphasized the oral component of the exam and how it reflects course goals.

In most Romance Language and Literature departments in the USA, an important objective is for students to be able to read authentic, literary texts before entering the literature course. Unfortunately, to date, it appears that no research has specifically examined variables involved in CBT and reading, even though [Benrhardt \(1996\)](#) expressed concern about the appropriacy of CBTs for assessing particular skills such as reading comprehension.¹ Would the inclusion of an additional factor such as self-assessment help program directors with decisions regarding placement of advanced learners? With advanced L2 learners of Spanish, the current research project attempts to determine the reliability of self-assessment as a predictor of achievement on the reading section of a CBT and as an indicator of performance in a subsequent class.

2. Review of the literature

2.1. Foreign language placement testing at the university level

In many university level institutions in the USA, students are required to take a foreign language placement exam for the following two reasons: (1) it guides students who will continue studying a foreign language already begun in high school to ascertain the appropriate level of course in which to enroll at the university, and (2) it allows students who do not wish to continue studying to test out of the either one or two-year language competency requirement. No placement exam is required of students who are beginning foreign language studies at the university; however, these students must enroll in the introductory level courses. The exams are also taken by currently enrolled students who have not taken any language for a few years and need a placement score prior to registration, and by transfer students in order to grant transfer credit.

Detailed information about foreign language placement exams is usually disseminated to all incoming freshmen. Instead of the traditional paper-and-pencil placement test, some universities offer the CBT. The CBT can be taken from any computer terminal at any time prior to arrival to campus. All students are given a username and password. The foreign language placement exam usually measures reading comprehension, vocabulary, grammar, listening, and in some cases, cultural knowledge. After all questions have been answered, results and placement are issued immediately. All in all, the main purpose of the placement exam is to provide information to help students enroll in courses that correspond to their levels of achievement. Some language program directors do not feel that the placement exam alone should determine correct placement. Given the wide range of instructional methodologies used in the USA at the high school levels, it is difficult to depend on years of prior language study for placement decisions and reading proficiency. The reliability of other factors, such as self-assessment, need to be considered for placement into upper level language courses.

In 1997, [Brown](#) contended that little is known about the use of computers in language testing and made a strong case for more research in this area. Even though there was a lack of research to support CBT, he specifically mentioned the following two immediate benefits that can be gained from computer-assisted language testing: “Computer-assisted language

¹ [Chalhoub-Deville \(2001\)](#) offers an excellent discussion of issues in computer-adaptive testing (CAT) of reading proficiency. The present study does not use CAT, and therefore this discussion is not included in the review of literature.

tests can be individually administered, even on a walk-in basis. Thus group-administered tests and all of the organizational constraints that they impose will no longer be necessary. Traditional time limits are not necessary. Students can be given as much time as they need to finish a given test because no human proctor needs to wait around for them to finish the test” (p. 45). More recently, [Bernhardt et al. \(2004\)](#) endorse these advantages of CBT.

2.2. Levels of foreign language instruction

The first and second year L2 courses in Romance Language departments in the USA generally utilize communicative language teaching methodologies that emphasize the development of speaking and listening skills. The reading of complete, authentic texts usually begins at the advanced levels of language instruction with courses on advanced grammar and composition. Consequently, the instructional practices shift from a focus on language skills to an emphasis on text analysis and interpretation. The objective is to prepare students for the level of reading and writing required in the literature and linguistic courses. By the time learners reach the advanced levels of language instruction, they should have completed approximately 700 class hours of exposure to the target language ([Lee, 1988](#)).

In most universities students in the advanced language courses enroll because they choose to, not because they are obliged to take the course in order to fulfill general language requirements. These courses are usually required for the major or minor. At this point we know very little about readers at the advanced levels of language instruction in empirical terms, and it is at this stage of acquisition where more L2 reading research is needed ([Brantmeier, 2001](#); [Young, 2003](#)). At the university where the present study was undertaken, the large number of freshmen test directly into the advanced level language courses. For example, in the Fall of 2004, approximately 100 students placed into the advanced intermediate or advanced level courses at University X. Even though students place into this level via the CBT, unfortunately, instructors often complain that students are not prepared to read the more lengthy, authentic texts, and consequently class time is spent deconstructing the plot instead of moving into textual analysis. The present study is motivated, in part, by the need to explore other factors that could aid program directors with decisions regarding placement of advanced learners.

2.3. Self-assessment about L2 learning abilities

Researchers ([Birckbichler et al., 1993](#); [Borrows et al., 1981](#); [Heilenman, 1991](#); [von Elek, 1987](#)) have found that students accurately make self-assessments about their ability and achievements in several disciplines. On the other hand, some researchers have found that less proficient learners overestimate their abilities ([Hilton et al., 1985](#); [Wesche et al., 1990](#)). More specifically, research findings suggest that people are generally overconfident on tests of acculturated knowledge, such as vocabulary and general knowledge tests ([Justlin, 1994](#); [Stankowv, 1998](#)).

Self-assessment, as used in second language learning contexts, is a technique utilized in some placement procedures that involves asking second language learners to make judgments about their own abilities ([Heilenman, 1991](#); [Schwartz, 1985](#)). [Oscarson \(1997\)](#) provides a review of investigations concerning self-assessment in the second and foreign language context and comments on previous reviews that exemplify the scarcity of research in this area ([Falchikov and Boud, 1989](#); [Heidt, 1979](#); [Oscarson, 1984](#)). Overall,

the studies that have been conducted generally support the use of self-assessment as an indicator of second language abilities.

Using a variety of rating instruments for self-assessment that included scaled descriptions of performance, [Oscarson \(1978\)](#) reported that adults studying EFL were able to make fairly accurate appraisals of linguistic abilities. The self-assessment instruments positively correlated with both teacher ratings and written test scores (see [Oscarson, 1997](#) for details). In a series of experiments with first-year students studying English and French as second languages at the University of Ottawa, [LeBlanc and Painchaud \(1985\)](#) found positive correlations between self-assessment questionnaires and standardized proficiency tests for listening and reading. They concluded that “self-assessment must be considered a very valuable tool as a placement instrument,” (p. 673). [Krausert \(1991\)](#) conducted an experiment to determine the usefulness of self-assessment for placement decisions with ESL students at the University of Southern California. Her findings indicated that university level ESL students do correctly assess their reading, writing, and speaking abilities, and she recommends that educators should utilize self-assessment instead of standardized exams for ESL placement in university programs. [Hargan \(1994\)](#) compared a traditional multiple choice placement test for grammar and a self-placement procedure with university students and found that both instruments indicated the same level of placement. [Birckbichler et al. \(1993\)](#) utilized a self-assessment questionnaire as part of the placement exam administered at Ohio State University in the USA and reported that self-assessment correlated higher than any other variable with the placement scores. Participants also reported accurate ability levels. [Ross \(1998\)](#) conducted a meta-analysis of experiential factors involved in self-assessment in L2 testing and found the largest number of correlations between L2 reading criterion variables and self-assessment in reading. Ross explains, “Reading tends to be the skill that is first taught in the foreign language context, and given the fact that most of the subjects were recruited from universities, subjects were most likely very experienced in using reading skills,” (p. 6). [Deville and Deville \(1999\)](#) suggest self-assessment as part of the procedures for estimating a starting point for computer-adaptive testing for L2 placement.² Most recently, with all students enrolled in advanced Spanish courses, [Brantmeier \(2005\)](#) found that levels of self-assessed abilities positively correlated with levels of enjoyment. The study also yielded significant effects for both self-assessed ability and enjoyment on written recall, but no such effects were found on multiple-choice questions. The above investigations lend support for the hypothesis that self-assessment can be accurate for placement. In a recent summary of research findings, [Oscarson \(1997\)](#) claimed that, “It seems to be fairly commonly agreed that the question of accuracy and appropriateness of self-estimates of proficiency depends, to a considerable degree, on feature of context and on the intended purpose of the assessment in each individual case” (p. 8). The specific query about whether self-assessment is an accurate and reliable predictor for language placement and classroom performance in reading with advanced learners in the USA has yet to be examined.

The present study borrows from the above studies and utilizes the term “self-assessment” to mean judgments or beliefs that learners make about their L2 learning abilities and performance, including assessment for placement ([Oscarson, 1980](#)).

² For an in-depth discussion of CAT and general language learning, see [Brown \(1997\)](#), [Dunkel \(1991\)](#), and [Chalhoub-Deville \(2001\)](#).

3. The present study

Unfortunately, to date, it appears that there is little empirical evidence concerning the association among L2 readers' self-assessment, CBT scores, and classroom based reading performance with learners at any level of language instruction. With learners from advanced levels of instruction, the present L2 reading study is a preliminary attempt to provide empirical evidence concerning the use of a self-assessment factor and CBT placement score as predictors of subsequent reading achievement. The following questions guide the present study:

1. Do self-assessment ratings of L2 reading ability accurately predict achievement on a computer-based placement test and subsequent reading performance?
2. Do self-assessment ratings of L2 reading classroom performance accurately reflect achievement?
3. Does performance on a computer-based placement test predict subsequent reading performance?

3.1. Participants

The Romance Languages and Literatures Department at University X first began using online language placement tests in the summer of 2001. The instrument, entitled the "Romance Languages and Literatures Online Placement Exam (RLL OPLE)", has been administered on a regular basis ever since, with the majority of the testing completed by freshmen during the summer months prior to arrival at the university. Immediately following the exam, scores are reported individually to students and later to their instructors.

Participants in the present study were 71 students, ages 19–22, enrolled in an advanced level Spanish grammar and composition course (Spanish 307). This third-year course was taught by five different instructors and was the first in a two-course sequence taken immediately before entering the literature courses. As part of the course students are assigned to read lengthy, authentic literary works from the literary canon. At the university where data was collected there is no language requirement, and therefore all students in the study enrolled in the course voluntarily.

To ensure a homogeneous population of participants, only students with the following criteria were included in the final data analysis: (1) students who achieved the appropriate composite score on the OPLE (tested into Spanish 307), (2) students whose native language was English, (3) students who enrolled in Spanish 307 the semester immediately after taking the OPLE; and (4) students who completed all tasks for both data collection settings. In the end, for this preliminary study, final data analysis included a total of 34 students (12 men and 22 women).

3.2. Self-assessment questionnaire

The self-assessment factors consisted of four questions, two completed before the OPLE and two completed after the in-class reading. All questions were created after reviewing prior questionnaires created for language placement contexts or language-related standardized exams administered in the USA. The 5-point Likert scale was used

to encourage more precision in rating and encouraged respondents to show greater discrimination in their judgments.

A questionnaire of self-assessment ratings for completion before learners read the passages was based on the instrument utilized by Birckbichler et al. (1993) and Deville and Deville (1999) for language placement at a university. The first question (Q1) about L2 reading read: “How well can you read in Spanish?” with five possible choices from 1 (not very well at all) to 5 (very well). In addition to this question, the following question was added (Q2): “How do you rate yourself as a reader of Spanish?” with five possible choices of 1 (I do not think that I am a good reader of Spanish at all) to 5 (I think I am an excellent reader of Spanish). See Appendix A for the self-assessment questionnaire utilized. The objective of pre-self-assessment items was to gain a general assessment. It is important to note that readers completed these questions online before beginning a timed placement examination that determined their level of Spanish, and this may be considered a high-stakes assessment situation. The self-assessment inventory also included the same question for listening, speaking, grammar, writing, and cultural knowledge; however, these items were not utilized in the present study that emphasizes reading.

A questionnaire of self-assessment ratings was also created 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. The first question was taken from a questionnaire created by Schraw et al. (1995) for an L1 reading investigation concerning sources of interest and effects on written recall. The question assessed readers’ ease of comprehension, and it read: “I found the passage I just read easy to understand” with five possible answers ranging from “I strongly disagree with this statement” to “I strongly agree with this statement.” This second question was based on follow-up items for self-assessment designed by Tobias and Everson (1998) and Everson (Chief Research Scientist, College Board of USA, personal communication, November, 2004). This question read: “How much of the passage did you understand?” with answers ranging from “I did not understand very much at all” to “I understood all of the passage”. See Appendix A.

3.3. Reading section for OPLE

The score from the reading section of the OPLE exam was used as one dependent variable for the present study. The reading section of the OPLE exam was originally created based on the ACTFL proficiency guidelines for Advanced and Advanced Plus readers.³ The reading section of the OPLE contains eight different vignettes of varying styles and

³ According to the ACTFL guidelines, Advanced Readers “. . . are able to read somewhat longer prose of several paragraphs in length, particularly if presented with a clear underlying structure. The prose is predominantly in familiar sentence patterns. Reader gets the main ideas and facts and misses some details. Comprehension derives not only from situational and subject matter knowledge but from increasing control of the language. Texts at this level include descriptions and narrations such as simple short stories, news items, bibliographical information, social notices, personal correspondence, routinized business letters, and simple technical material written for the general reader”. Advanced Plus readers should be “. . . able to follow essential points of written discourse at a superior level in areas of special interest or knowledge. They are able to understand parts of texts which are conceptually abstract and linguistically complex, and/or texts which treat unfamiliar topics and situations, as well as some texts which involve aspects of target-language culture. They are able to comprehend the facts to make appropriate inferences. An emerging awareness of the aesthetic properties of language and of its literary styles permits comprehension of a wider variety of texts, including literary”.

lengths. The readings include excerpts about the daily lives of students, historical vignettes, a poem, personal narratives, and encyclopedia-like readings. Comprehension is measured via multiple choice items. For each of the multiple choice questions four possible responses were created: one correct response and three distractors. All distractors in the multiple-choice questions were plausible (Alderson, 2000; Bernhardt, 1991; Wolf, 1993), and all were written in Spanish. Some questions included inferential items as well.

3.4. Reading passage for classroom performance

The reading passage was selected after carefully examining reading materials used for this level of language instruction. The short story used in this study, *Aniversario*, by Luis Romero, was taken from an anthology for advanced readers entitled *Aproximaciones al estudio de la literatura hispánica*, by Virgilio, Friedman, and Valdivieso. The story consists of 1270 words and was kept in its original form, including word glosses. Investigations concerning the role of schemata in L2 reading comprehension have revealed that what students already know (the knowledge stored in their existing cognitive domain) significantly affects their comprehension of L2 reading materials (Carrell, 1983; Hudson, 1982; James, 1987; Johnson, 1981). Other studies found significant gender differences with topic familiarity levels and comprehension (Bügel and Buunk, 1996; Brantmeier, 2002, 2003; Schueller, 1999). Reading researchers need to take both passage content and gender into account when conducting investigations about the L2 reading process (Chavez, 2001), and therefore the present study went to great lengths to control for gender differences in topic familiarity. Before the investigation, the reading passage for the present study was piloted with 67 students to ensure that it contained a topic familiar to students. The 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

Development of test design and procedures has long been investigated in reading research, and Alderson (2000) argues that there is not one best method for testing reading. Common assessment measures used by L2 reading researchers include multiple choice, written and oral recall, cloze test, sentence completion, open ended questions, true/false, matching activities, checklists, fill in the blanks, etc. The outcome of each individual assessment task provides a limited representation of reading comprehension, and to understand the complete picture and to be able to generalize, a variety of assessment tasks are needed to measure comprehension (Bernhardt, 1991). To obtain a more complete representation of comprehension in the present study, three different assessment tasks were utilized: written recall, sentence completion, and multiple-choice.

The free written recall has been utilized to assess L2 reading comprehension for more than two decades (Bernhardt, 1983; Brantmeier, 2002; Carrell, 1983; Lee, 1986; Young and Oxford, 1997; and more). This written recall protocol does not influence a reader’s understanding of the text (Bernhardt, 1991) as there is no tester interference and there are no retrieval cues provided. For the present study, immediately after reading the passage, the written recall protocol asked readers, without looking back at the passage, to recall and write down as much as they could of what they just read.

Most recall tasks are scored by tallying the quantity of correct information recalled with a variety of scoring rubrics including idea units, propositional units, pausal units, etc. (Barnett, 1988; Brantmeier, 2002; Carrell, 1983; Lee, 1986a,b; among others). The present study utilizes the pausal unit protocol to analyze the text and recalls. A pausal unit is a unit or entity that during normally paced oral reading has a pause on each end of it (Bernhardt, 1991). For the present study, pausal units were not ranked in terms of salience to the message of the text. The present investigation does not examine weighted pausal units but rather the total number of correct pausal units recalled from the test. The pausal unit was utilized for practical purposes as the present study was conducted for program placement objectives. Alderson (2000) cites Bernhardt's (1991) study where she reports that the Meyer's (1975) recall scoring protocol can take between 25 and 50 hours to develop one template for a 250 word text. The present study utilized a 1270 word text, and the intention of the study is to aid those in charge of language placement. Consequently, the pausal unit protocol was followed. For the present study, oral reading by four native speakers was used for the initial division of pausal units. The units were listed, and then written recall protocols were checked for the presence of absence of such units. The total number of pausal units was utilized for the written recall score.

Sentence completion is an open response task where certain limits are placed on possible answers. The sentences are created so that all possible answers are foreseeable, and the objectivity of scoring depends upon the comprehensiveness of the answer key. Multiple choice questions are the most widely used means of assessing reading comprehension. The answers are pre-determined and there is no ambiguity in the scoring of right/wrong answers. For each of the multiple choice questions four possible responses were created following the same rubric as the multiple choice items on the OPLE (Alderson, 2000; Bernhardt, 1991; Wolf, 1993). For the multiple choice items in this part of the study, another condition was also added: the test-takers were not able to determine correct responses by looking at the other questions on the page. Each sentence completion item corresponded with a multiple choice question (solicited the same information).

Language of questions is also of great concern to test constructors. To avoid a test of writing instead of reading, L2 reading researchers also recommend that the recall be written in the reader's native language (Alderson, 2000; Bernhardt, 1991; Lee and Ballman, 1987; Wolf, 1993). In the present study all assessment tasks for subsequent reading performance were completed in the reader's native language.

3.6. *Topic familiarity*

Participants completed a questionnaire that included information such as sex, age, major, native language, and number of years of Spanish study in high school and university. 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*

All incoming freshmen were required to take the OPLE, and these scores were used to determine placement. Before completing the OPLE exam, all students were asked to assess

their L2 reading abilities. These ratings were used for the self-assessment factor. In order to assess 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: reading passage, self-assessment questionnaire, written recall, sentence completion items, multiple-choice questions, and topic familiarity questionnaire. Before the experiment participants were invited to participate in an experiment concerning L2 reading, and they were also told that they would complete the above tasks. No participants declined to participate in the study. The researcher and instructors for the courses were present during all data collection sessions so that students would not look back at any previous pages while reading and completing all tasks.

3.8. Data analysis

The self-assessment rating was the independent variable. The total number of correct responses was the dependent variable used for the reading section of OPLE. Subsequent reading performance consisted of three different dependent variables: 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 0.96. In the end, the total number of pausal units was 133. For sentence completion items there were several possible correct answers for each question. Separately the raters identified all possible correct responses, and the interrater reliability index was 0.94. For multiple choice, there was one correct answer for each question.

Means and standard deviations were calculated for each research question. In an attempt to predict the OPLE reading score and subsequent reading performance from the self-assessment scores, data were examined using regression analysis. The regression procedure examined the strength of these relationships as well as the amount of variance explained by self-assessment. The Alpha level for statistical significance was set at 0.05.

4. Results

The following are the mean scores and standard deviations for all four self-assessment items: Pre-question 1 “*How well can you read in Spanish?*” ($M = 3.4$, $SD = 0.74$) = “*Ok*”; Pre-Question 2 “*How do you rate yourself as a reader of Spanish?*” ($M = 3.0$, $SD = 0.52$) = “*I am an Ok reader of Spanish*”; Post-Question 3 “*I found the passage I just read easy to understand*” ($M = 2.9$, $SD = 0.85$) = “*I somewhat agree with this statement*”; and, Post-Question 4 “*How much of the passage did you understand?*” ($M = 2.8$, $SD = 0.67$) = “*I understood some of the passage*” Fig. 1 graphically depicts mean scores of each item for self-assessment of L2 reading ability, and Fig. 2 offers a graphic depiction of frequency scores for each self-assessment item. For Pre Q1, only one student indicated the highest ranking of “I read very well in Spanish”, and only one student indicated the lowest ranking of “I do not read very well in Spanish at all”. For Pre Q2, no student indicated the highest and lowest ranking of “I think I am an excellent reader of Spanish” or “I do not think that I am a good reader of Spanish at all”. For the first post-reading self-assessment item, two participants reported that they “strongly disagree” and no students “strongly agreed” with the statement about the passage being easy to understand. For

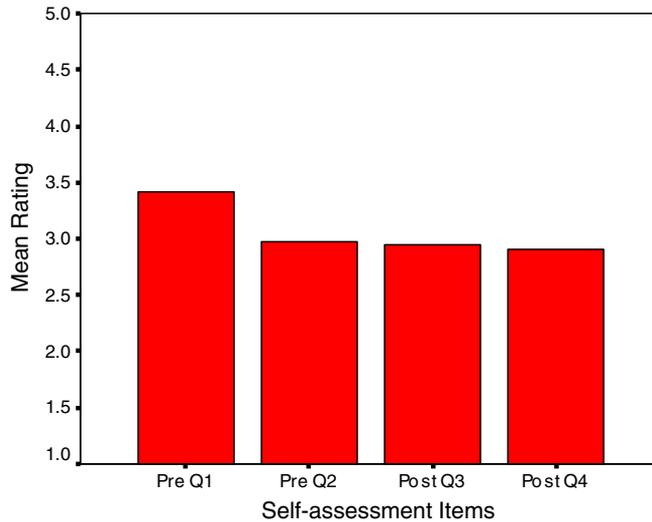


Fig. 1. Self-assessment of L2 reading ability: mean scores for four items; Pre Q1 = How well can you read in Spanish? Pre Q2 = How do you rate yourself as a reader of Spanish? Post Q3 = I found the passage I just read easy to understand. Post Q4 = How much of the passage did you understand?

Post Q4 only one student reported not understanding very much at all of the passage, while no students reported understanding the entire passage.

Means scores and standard deviations for all comprehension assessment tasks are listed on Table 1. As listed on the table, participants scored a mean score of 19.7, or an average of 66% on the OPLE reading test. Fig. 3 depicts the distribution of OPLE reading scores for the advanced learners. The mean score of the written recall was extremely low with participants only recalling a mean score of 12.9, or 10% of the total possible pausal units. On sentence completion items, participants scored a mean score of 6.8 or 68%, and for multiple choice questions, participants scored a mean score of 7.8, or 78%.

As indicated in Table 2, results of the regression analysis revealed no significant associations between reader's self-assessment (Q1 and Q2) and OPLE reading score, and reader's self-assessment (Q1 and Q2) and subsequent reading performance as measured via recall, sentence completion, and multiple choice. Likewise, no significant associations were found between post-self-assessment items and all three measures of comprehension. A regression analysis was conducted to examine the association between OPLE reading scores and subsequent classroom reading achievement, and findings yielded a significant association between the following: OPLE score and recall score, and OPLE score and MC score. No significant association was found between OPLE score and SC.

In order to graphically depict the strength of the significant relationship between OPLE reading scores and comprehension and recall scores and OPLE reading scores and multiple choice scores, all possible pairs of values are shown on scatterplots. Fig. 4 is a scatterplot of OPLE score by Recall. OPLE reading scores are on the vertical axis with a range of 1–27, and Recall scores are on the horizontal axis with a range of 2–30. Further analysis were conducted to check for linear and/or curvilinear trends between OPLE reading scores and Recall scores, and results indicated that both trends were statistically significant

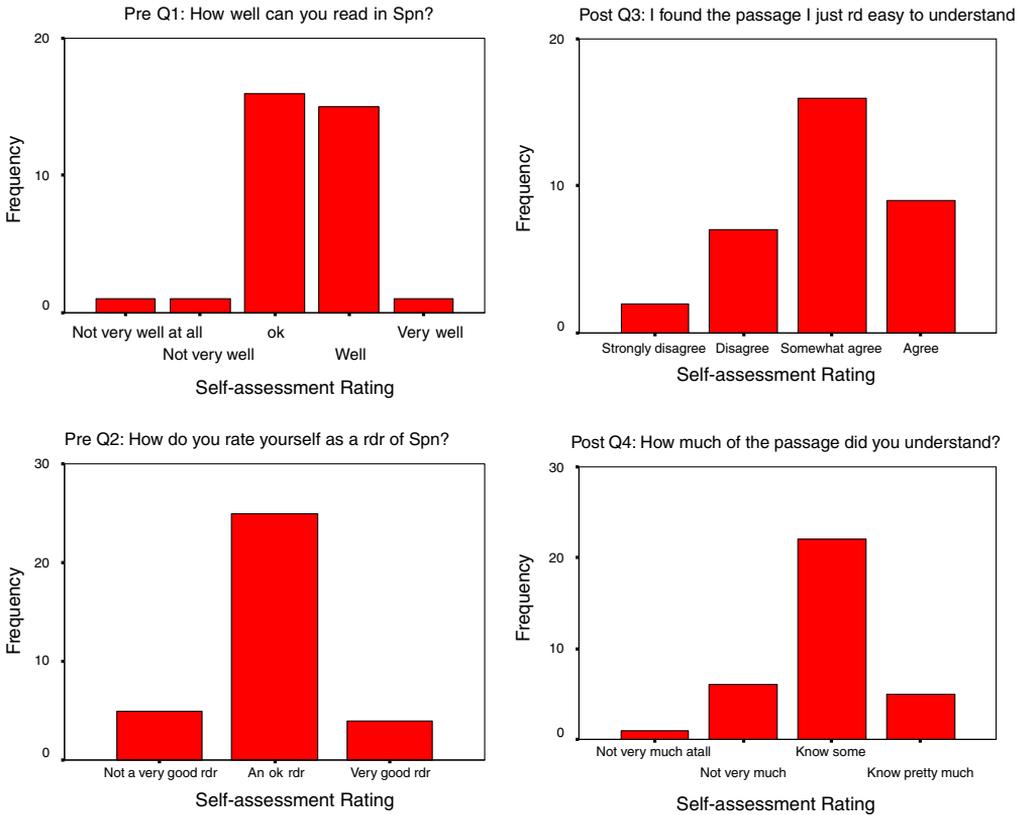


Fig. 2. Self-assessment rating.

Table 1

Mean scores and standard deviations for OPLE reading and subsequent achievement

	OPLE reading	Recall	SC	MC
Possible score	30	133	10	10
Mean	19.7	12.9	6.8	7.8
SD	6.6	7.3	2.1	1.7
Percentage (score/#possible)	66%	10%	68%	78%
Minimum	1	2	2	5
Maximum	27	30	10	10
Range	26	28	8	5

n = 34.

(linear = 0.01; curvilinear = 0.02). Fig. 5 is a scatterplot of OPLE score by multiple choice. OPLE reading scores are on the vertical axis with a range of 1–27, and multiple choice scores are on the horizontal axis with a range of 5–10. Visual inspection of this figure also revealed a possible linear and curvilinear trend between OPLE reading scores and multiple choice scores, and further analysis indicated that both relationships are statistically significant (linear = 0.00; curvilinear = 0.00).

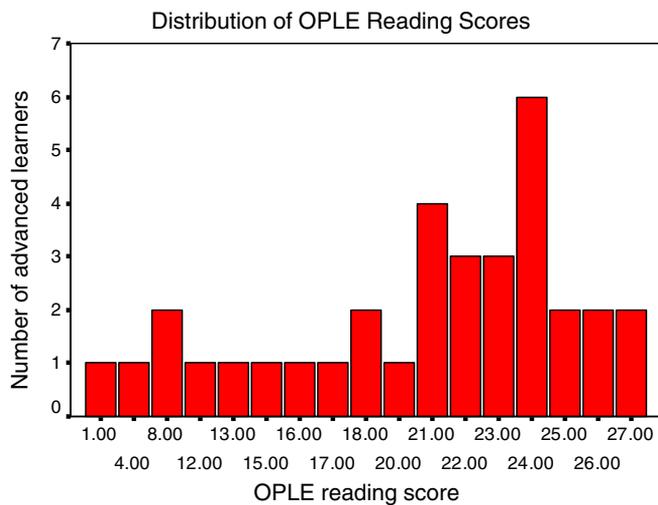


Fig. 3. Distribution of OPLE reading scores with advanced language learners.

Table 2

Regression analysis

Predictors (constant)	R^2	T -ratio	P
<i>Reader's self-assessment: pre-reading</i>			
Q1			
OPLE	0.00	3.72	0.90
Recall	0.01	1.60	0.59
SC	0.01	3.29	0.52
MC	0.02	4.95	0.40
Q2			
OPLE	0.04	1.79	0.24
Recall	0.11	0.09	0.06
SC	0.04	2.16	0.29
MC	0.10	3.06	0.07
<i>Reader's self-assessment: post-reading</i>			
Q3			
OPLE	0.01	4.26	0.64
Recall	0.01	3.11	0.74
SC	0.03	4.26	0.36
MC	0.03	6.67	0.34
Q4			
OPLE	0.01	3.30	0.59
Recall	0.01	1.64	0.54
SC	0.11	2.40	0.06
MC	0.02	5.37	0.48
<i>OPLE reading score</i>			
Recall	0.20	0.86	0.01*
SC	0.03	5.01	0.37
MC	0.32	6.67	0.00*

* $P < 0.05$.

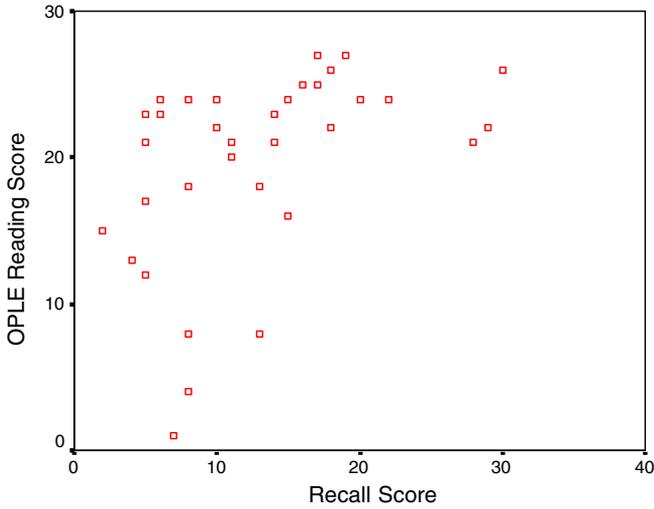


Fig. 4. Scatterplot of OPLE reading score by recall score.

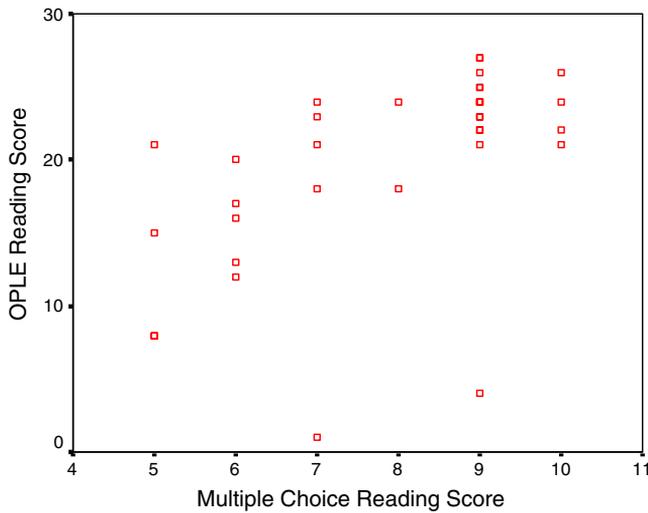


Fig. 5. Scatterplot of OPLE reading score by multiple choice reading score.

5. Discussion and implications

The mean score for the first self-assessment question was slightly higher than that of the other three questions with students generally indicating that they are “ok” readers of Spanish. However, the number of instances within a particular rating varies (Fig. 2 offers a visual display of frequency for each rating). An examination of how many students rated themselves at each point on the 5-point scale shows that there is a certain degree of variation within the group of advanced language learners. Not all learners who tested into Advanced Spanish rated themselves the same for reading, and Fig. 3 shows the great var-

iation of OPLE reading scores for this same group of advanced learners. These individual differences in self-assessment are important to consider when interpreting findings concerning the association of self-assessment ratings and reading achievement. A close look at the bar graphs reveals that different patterns emerged with Pre-reading Question 1 and Pre-reading Question 2, but the same patterns emerged for both Post-reading questions. Further analysis indicated that all four self-assessment items were significantly correlated.

There was not a strong relationship between all four self-assessment items and reading achievement (OPLE scores and all three subsequent comprehensions tasks). These findings are confusing because the mean self-assessment scores indicate “Ok” readers, and average performance on the OPLE and sentence completion could be interpreted as “Ok” (OPLE = 66%; multiple choice = 78%). One way to interpret this would be, again, to highlight the degree of individual differences in both self-assessment ratings as well as performance on the OPLE and multiple choice items. The average score for recall was only 10%, which may indicate that the written recall test was too difficult for the reading passage. Furthermore, the association between the OPLE and recall measures was weak ($r^2 = 0.20$). Perhaps with longer passages (approximately 1200 words) the written recall should not be utilized as it could be insensitive to real differences between participants. As discussed earlier, for practical purposes of placement, the written recall was not scored according to weighted pausal units and this decision could have influenced results. If a future study of this nature utilizes the written recall protocol, perhaps weighted pausal units should be tallied and examined.

Overall, these results contradict earlier findings with language learners where self-assessment positively correlated with placement scores, ability and achievement (Birckbichler et al., 1993; Brantmeier, 2005; Hargan, 1994; Heilenman, 1991; Krausert, 1991; Oscarson, 1978; Ross, 1998; von Elek, 1987). It is important to note that Brantmeier (2005) utilized all learners enrolled in Advanced Spanish. The present study utilizes only freshmen who had taken an online placement exam and placed directly into the advanced courses, and this could help explain the contradictory findings with advanced learners. Results of the present study echo other investigations where learners do not accurately assess their abilities (Hilton et al., 1985; Wesche et al., 1990). These inconsistent findings underscore the need for more research that explores self-assessment as a factor to be considered specifically for advanced language placement.

The construct of self-assessment determined by the questionnaire utilized in the present study may be too limited and narrow for advanced readers. LeBlanc and Painchaud (1985) contend that self-assessment questionnaires should give learners the opportunity to indicate what they think they can do with the language they are studying. Perhaps if the self-assessment inventory was worded differently, data would yield different results. For example, the *Reading Skills Self-Assessment Grid* is a component of the *European Language Portfolio* developed under the direction of the European Council. It is used to develop a self-assessment of reading skills in English as a Foreign Language. This questionnaire asks learners which description most closely matches their comprehension level when reading an English text. This 6-point rating scale ranges 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”. These ratings are utilized for a CAT (computer-adaptive testing)

situation as a starting point, but they could also be modified and utilized for a self-assessment factor to help with general placement decisions. Oscarson's (1980) self-assessment questionnaire includes a 5-point scale with specific details about learner abilities such as "I understand everything or nearly everything written in the language in non-specified fields. There may be words I do not understand in difficult texts" and "I cannot read the language at all". These more detailed items could be modified to match language program levels (see Oscarson, 1978 for a more detailed description of types of questions). LeBlanc and Painchaud (1985) developed a questionnaire with a 5-point scaled that included statements graded for difficulty level, such as "I understand short and simple written communications (posters, schedules, announcements)" and "I read specialized articles concerning my fields of study fluently". A future study could utilize the same data collection instruments and procedures as the present study with more descriptive self-assessment items such as the above. A possible question could be, "I understand short stories written by native speakers". Furthermore, a future study could ask readers in more detail about how well they think they performed on a particular reading in addition to asking them to rate their overall reading abilities. In the present study, the inventory of overall reading ability was borrowed from previous studies and the pre-reading self-assessment questions did not require a lot of time to complete. The self-assessment inventory, if used for placement, would be administered online before completing the actual placement exam. The issue of time is important to consider. LeBlanc and Painchaud (1985) reported that their self-assessment inventory takes about 20 min to complete. Perhaps the inventory could be modified and shortened to meet the learner's potential needs as a reader at the advanced level.

A closer look at the self-assessment and comprehension scores for subsequent classroom performance reinforces the need for more research on self-assessment as an indicator of reading ability at the advanced levels. Generally, the emphasis in the early years of language instruction in the USA is on oral, communicative skills and reading is often a silent task completed at home. Perhaps before students enter the university, reading comprehension is a skill that has never been effectively evaluated before, and therefore readers have no foundation for self-assessment. Students may not have regularly monitored their L2 reading progress at the secondary levels, and they may not have had previous feedback on their L2 reading achievement. In other words, students may likely not fully understand what it means to be an "excellent" or "ok" reader of Spanish. Entering freshmen may not be equipped with the skills necessary to adequately evaluate their L2 reading abilities, which reinforces the need for more contextualized, criterion-referenced self-assessment instrument for L2 reading. Moreover, Oscarson (1997) claims that self-assessment procedures seem to be useful in day-to-day teaching/learning contexts, whereas the use of self-assessment for "high-stakes" purposes appears to be inappropriate. Adult learners may need training to be able to make reliable judgments on their abilities. Self-assessment and self-monitoring could be stressed in teacher training courses and workshops (Oscarson, 1997) so that at the university level students are accustomed to self-evaluation.

A regression analysis revealed a significant association between the following: OPLE score and recall score, and OPLE score and multiple choice score. Further analysis revealed significant linear and curvilinear trends. These results attest to the reliability of the OPLE as a predictor of subsequent performance, as students performed equally well on the reading part of the OPLE and subsequent classroom testing. The OPLE relies on a structured response format to assess reading comprehension, and subsequent perfor-

mance was measured via an open-ended format and the multiple choice items. Scores on both assessment task formats were significantly related to prior scores on OPLE, which underscores the fact that the OPLE can be used to predict actual reading performance in the course. Unfortunately, what this also suggests is that learners testing into the advanced levels may not read at the appropriate level. The OPLE exam has three sections, and results show that students scored significantly higher on both the grammar and listening sections than on the reading section, and consequently test takers achieve the appropriate overall composite score necessary to enter the advanced language courses. However, results of the present study suggest that these advanced language learners are not reading at an advanced or advanced plus level.⁴ Findings also show the great variation of reading abilities within this group of advanced learners.

Given that students are not reading at the appropriate level and that they may not be equipped with the skills to self-evaluate reading ability, instructors at the advanced levels could guide readers in the development of self-assessment methods through in-class discussions, teacher–student interviews, and weekly self-evaluations including self-assessment checklists and inventories. Instructors could require students to review their work and evaluate their progress, and consequently students may develop the ability to understand the criteria for successful L2 reading and consequently may more adequately self-evaluate. A future study could examine instructional practices that enhance self-assessment of reading abilities, and also explore whether these practices enhance reader performance at this level. Perhaps if students are better equipped to evaluate themselves after months of in-class self-evaluation activities, they would achieve higher reading scores. The present study is not an attempt to examine metacognitive skills for reading, and it did not ask students to complete the OPLE exam at the end of the course to see if they would do significantly better on the second administration of the test. At this point it is too early to make assertions concerning these issues.

6. Conclusion

Language testing for placement remains a complex activity at universities in the USA. By examining the relationship between learner's self-assessment and scores on an online reading placement test as well as subsequent reading performance, the present study serves as a preliminary attempt to test the reliability of self-assessment as a factor to be used for predicting placement with advanced learners. Findings show that we cannot depend on self-assessment of reading ability, as measured via non-descriptive self-assessment items, with advanced learners who are entering the university. Results highlight the need for a descriptive and criterion-referenced self-assessment instrument for use with advanced placement in reading. Self-assessment with online placement testing can bring together a number of related disciplines (Everson, personal communication, November, 2004), and Bernhardt (2005) echoes this suggestion by stating that L2 reading is a crucial area of research that should transcend the borders of applied linguistics (p. 133). Future studies could modify existing self-assessment instruments utilized in Europe and from related disciplines to test their validity for second language placement. Surprisingly, results of the present study also indicate that learners at the advanced level of language instruction

⁴ The advanced and advanced plus levels refer to the guidelines set forth by ACTFL.

may not be reading at an appropriate level, and this is another issue worthy of future inquiry. Overall, findings underscore the need for more investigations concerning factors involved in reading and foreign language placement with advanced language learners.

Appendix A

A.1. Pre-reading self-assessment questionnaire

Please circle the correct answer.

1. How well can you SPEAK Spanish?
 - (a) Not very well at all.
 - (b) Not very well.
 - (c) Ok.
 - (d) Well.
 - (e) Very well.

2. How well can you UNDERSTAND Spanish?
 - (a) Not very well at all.
 - (b) Not very well.
 - (c) Ok.
 - (d) Well.
 - (e) Very well.

3. How well can you READ Spanish?
 - (a) Not very well at all.
 - (b) Not very well.
 - (c) Ok.
 - (d) Well.
 - (e) Very well.

4. How well can you WRITE Spanish?
 - (a) Not very well at all.
 - (b) Not very well.
 - (c) Ok.
 - (d) Well.
 - (e) Very well.

5. How much do you know about the CULTURE of the Spanish speaking countries?
 - (a) Not very much at all.
 - (b) Not very much.
 - (c) Know some.
 - (d) Know pretty much.
 - (e) Know a lot.

6. How do you rate yourself as a reader of Spanish?
 - (a) I am not a good reader of Spanish.

- (b) I am not a very good reader of Spanish.
- (c) I am an OK reader of Spanish.
- (d) I am a very good reader of Spanish.
- (e) I am an excellent reader of Spanish.

Self-assessment questionnaire taken from Birckbichler et al. (1993) and Deville and Deville (1999).

A.2. Post-reading self-assessment questionnaire

1. I found the passage I just read easy to understand.
 - (a) I strongly disagree with this statement.
 - (b) I disagree with this statement.
 - (c) I somewhat agree with this statement.
 - (d) I agree with this statement.
 - (e) I strongly agree with this statement.

Schraw et al. (1995).

2. How much of the passage did you understand?
 - (a) I did not understand very much at all.
 - (b) I did not understand very much.
 - (c) I understood some of it.
 - (d) I understood most of it.
 - (e) I understood all of it.

Tobias and Everson (1998).

References

- Alderson, C.A., 2000. *Assessing Reading*. Cambridge University Press, United Kingdom.
- Bachman, L.F., 1990. *Fundamental Considerations in Language Testing*. Oxford University Press, Oxford.
- 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.
- Bennhardt, E.B., 1996. If reading is reader-based, can there be a computer adaptive reading test? In: Chalhoub-Deville, M. (Chair), *Issues in Computer-adaptive Testing of Second Language Reading Proficiency*. Center for Advanced Research on Language Acquisition of the University of Minnesota, p. 18.
- Bernhardt, E.B., 2005. Progress and procrastination in second language reading. *Annual Review of Applied Linguistics* 25, 133–150.
- Bernhardt, E.B., Rivera, R.J., Kamil, M., 2004. The practicality and efficiency of web-based placement testing for college-level language programs. *Foreign Language Annals* 37 (3), 356–366.
- Birckbichler, D., Corl, K., Deville, C., 1993. *The dynamics of language program testing: implications for articulation and program revision*. The Dynamics of Language Program Direction. Heinle & Heinle, Boston, MA.
- Borrows, Ager, T.S.M., Bennette, M.F., Braun, H.I., Clark, J.L.D., Harris, L.B., Klein, S.F., 1981. *College students' knowledge and beliefs: a survey of global understanding*. The Final Report of the Global Understanding Project. Change Press, New Rochelle, NY, ERIC ED 215 653.

- Brantmeier, C., 2001. Second language reading research on passage content and gender: challenges for the intermediate level curriculum. *Foreign Language Annals* (34), 325–333.
- 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., 2005. Non-Linguistic variables in advanced L2 reading: learner's self-assessment and enjoyment. *Foreign Language Annals* 38 (4), 493–503.
- Brown, J.D., 1997. Computers in language testing: present research and some future directions. *Language Learning and Technology* 1 (1), 44–59.
- 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.
- Byrnes, H., 1991. In: Birckbichler, D.W. (Ed.), *Issues in Foreign Language Program Articulation*. Critical Issues in Foreign Language Instruction. Heinle & Heinle, Boston, pp. 6–28.
- 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.
- Chalhoub-Deville, 2001. Language testing and technology: past and future. *Language Learning and Technology* 5 (2), 95–98.
- Chapelle, C., 2001. *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing and Research*. Cambridge University Press, Cambridge.
- Chavez, M., 2001. *Gender in the Language Classroom*. McGraw-Hill, New York.
- Deville, M., Deville, C., 1999. Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273–299.
- Dunkel, A., 1991. Research on the effectiveness of computer-assisted instruction and computer-assisted language learning. In: Dunkel, P. (Ed.), *Computer-assisted Language Learning and Testing*. Newbury House, New York, pp. 5–36.
- European Language Portfolio Self-Assessment. Available from: <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.
- 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.
- Heilenman, K., 1991. Self-assessment and placement: a review of the issues. In: Teschner, R.V. (Ed.), *Assessing Foreign Language Proficiency of Undergraduates*, AAUSC Issues in Language Program Direction. Heinle & Heinle, Boston, pp. 93–114.
- Hilton, T., Grandy, J., Green Kline, R., Liskin-Gasparro, J.E., 1985. *Final Report: The Oral Language Proficiency of Teachers in the United States in the 1980s – An Empirical Study*. Educational Testing Service, Princeton, NJ.
- Hudson, T., 1982. The effects of induced schemata on the 'short circuit' in L2 reading: non-decoding factors in L2 reading performance. *Language Learning* 32 (1), 1–31.
- James, D., 1987. *Bypassing the Traditional Leadership: Who's Minding the Store? Profession*. Routledge, New York, pp. 41–53.
- Johnson, P., 1981. Effects on comprehension of building background knowledge. *TESOL Quarterly* 16, 503–516.
- Justin, P., 1994. The overconfidence phenomenon as a consequence of formal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes* 57, 226–246.
- Krausert, S.R., 1991. Determining the usefulness of self-assessment of foreign language skills: post-secondary ESL students' placement contribution. Ph.D. Diss., University of Southern California.
- Lange, D.L., Prior, P., Sims, W., 1992. Prior instruction, equivalency formulas, and functional proficiency: examining the problem of secondary school–college articulation. *The Modern Language Journal* 76, 284–294.
- 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., Dvorkin, T.R., Lee, J.F. (Eds.), *Foreign Language Learning: A Research Perspective*. Newbury House, Rowley, MA, pp. 108–117.
- Meyer, B.F.J., 1975. Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior* 7, 7–47.
- Oscarson, M., 1978. *Approaches to Self-assessment in Foreign Language Learning*. Council of Europe, Council for Cultural Cooperation, Strasbourg.
- Oscarson, M., 1980. *Approaches to Self-Assessment in Foreign Language Learning*. Pergamon Press, Oxford.
- 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., 1997. Self-assessment of foreign and second language proficiency. *The Encyclopedia of Language and Education*, vol. 7. Academic Publishers, Kluwer, pp. 175–187.
- Ross, S., 1998. Self-assessment in second language testing: A meta-analysis and 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.
- Schwartz, L., 1985. The California study on foreign language articulation. *ADFL Bulletin* 17 (1), 22–28.
- Stankowicz, L., 1998. Calibration curves, scatterplots, and the distinction between general knowledge and perceptual tests. *Learning and Individual Differences* 8, 28–51.
- Tobias, S., Everson, H.T., 1998. 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.
- Virgilio, C., Friedman, E., Valdivieso, T., 1998. *Aproximaciones al Estudio de la Literatura*. McGraw-Hill, Boston.
- von Elek, T., 1987. A test of Swedish as a second language: an experiment in self-assessment. In: Lee, Y., Fok, A., Lord, R., Low, G. (Eds.), *New Directions in Language Testing*. Oxford University Press, Oxford, pp. 47–57.
- Wesche, M.B., Morrison, F., Ready, D., Pawley, C., 1990. French immersion: post-secondary consequences for individuals and universities. *Canadian Modern Language Review* 46, 430–451.
- Wheritt, I., Clearly, T.A., 1990. A national survey of Spanish language testing for placement and outcome assessment at B.A.-granting institutions in the United States. *Foreign Language Annals* 23, 157–165.
- Wolf, D., 1993. A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal* 77, 473–489.
- Young, D.J., 2003. *Workshop on overcoming anxiety in the foreign language classroom* Language Learning Series. Washington University in St. Louis.
- 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.