The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals

Viorica Marian
Henrike K. Blumenfeld
Margarita Kaushanskaya
Northwestern University, Evanston, IL

Purpose: To develop a reliable and valid questionnaire of bilingual language status with predictable relationships between self-reported and behavioral measures.
Method: In Study 1, the internal validity of the Language Experience and Proficiency Questionnaire (LEAP-Q) was established on the basis of self-reported data from 52 multilingual adult participants. In Study 2, criterion-based validity was established on the basis of standardized language tests and self-reported measures from 50 adult Spanish–English bilinguals. Reliability and validity of the questionnaire were established on healthy adults whose literacy levels were equivalent to that of someone with a high school education or higher.
Results: Factor analyses revealed consistent factors across both studies and suggested that the LEAP-Q was internally valid. Multiple regression and correlation analyses established criterion-based validity and suggested that self-reports were reliable indicators of language performance. Self-reported reading proficiency was a more accurate predictor of first-language performance, and self-reported speaking proficiency was a more accurate predictor of second-language performance. Although global measures of self-reported proficiency were generally predictive of language ability, deriving a precise estimate of performance on a particular task required that specific aspects of language history be taken into account.
Conclusion: The LEAP-Q is a valid, reliable, and efficient tool for assessing the language profiles of multilingual, neurologically intact adult populations in research settings.

KEY WORDS: bilingualism, self-assessment, second language, language proficiency, questionnaire development

Bilingualism and multilingualism are the norm rather than the exception in today's world (Harris & McGhee-Nelson, 1992), and the proportion of linguistically diverse populations is increasing in the United States (U.S. Bureau of the Census, 2003). These demographic changes are reflected in the growing representation of multilingual and multicultural populations in research and applied settings. However, research with bilinguals often yields inconsistent findings (e.g., Grosjean, 2004; Marian, in press; Romaine, 1995). For example, bilingual cortical organization (e.g., Kim, Relkin, Lee, & Hirsch, 1997; Marian, Spivey, & Hirsch, 2003; Perani et al., 1998; Vaid & Hull, 2002), lexical processing (e.g., Chapnik-Smith, 1997; Chen, 1992; Kroll & de Groot, 1997), and phonological and orthographic processing (e.g., Doctor & Klein, 1992; Grainger, 1993; Macnamara & Kushnir, 1971; Marian & Spivey, 2003) have all been found to differ

1092-4388/07/5004-0940

depending on bilinguals' ages of language acquisition, mode(s) of acquisition, history of use, and degree of proficiency and dominance. These inconsistencies are further exacerbated by the absence of uniform assessment instruments in bilingualism research. Those who work with bilinguals and multilinguals often face the challenge of testing individuals whose language they do not speak (Roseberry-McKibbin, Brice, & O'Hanlon, 2005) and thus have to rely exclusively on self-assessed information, usually collected with improvised questionnaires. The need for a language self-assessment tool that is comprehensive, valid, and reliable across bilingual populations and settings prompted a systematic approach to developing the present Language Experience and Proficiency Questionnaire (LEAP-Q; see Appendix).

Previous Self-Assessment Studies

In general, previous research suggests that self-reported language measures are indicative of linguistic ability (e.g., Bachman & Palmar, 1985; MacIntyre, Noels, & Clement, 1997; Ross, 1998; Shameem, 1998; Stefani, 1994). Existing self-assessment tools for studying bilinguals span both domain-general (e.g., Bahrick, Hall, Goggin, Bahrick, & Berger, 1994; Delgado, Guerrero, Goggin, & Ellis, 1999) and domain-specific proficiency (e.g., Flege, Yeni-Komishian, & Liu, 1999; Jia, Aaronson, & Wu, 2002; Vaid & Menon, 2000). For instance, in a study of the relationship between self-reported proficiency and language performance, Delgado et al. (1999) tested Spanish–English bilinguals and correlated self-assessed proficiency in English and Spanish with performance on the Woodcock–Muñoz Language Survey (Woodcock & Muñoz-Sandoval, 1993). Delgado et al. found that participants assessed first-language (L1) skills more accurately than they did second-language (L2) skills. Woodcock–Muñoz scores correlated with all self-reported measures of L1 proficiency but with only self-reported measures of L2 reading and writing (and not with L2 speaking and understanding). Similarly, Bahrick et al. (1994) found that language dominance ratings correlated highly with performance on some tasks (e.g., category generation and vocabulary recognition) but correlated less with performance on other tasks (e.g., oral comprehension). Together, studies of domain-general self-assesement in bilinguals suggest that the relationship between self-reported and behavioral measures of language performance varies across languages and tasks (e.g., Bahrick et al., 1994; Delgado et al., 1999).

Studies of domain-specific proficiency assessment in bilinguals have focused on grammatical ability (e.g., Jia et al., 2002), the degree of foreign accent (e.g., Flege et al., 1999), and computational language use (e.g., Vaid & Menon, 2000). For instance, Jia et al. used a 32-item questionnaire that assessed the following four areas: (a) age/time variables associated with L2 acquisition; (b) environmental variables (e.g., number of L2 speakers at home; frequency with which L2 is spoken at home and in the workplace; and father's, mother's, and siblings' proficiency in speaking, reading, and writing L2); (c) affective variables (e.g., self-consiousness, cultural preference and identity, and motivation); and (d) self-evaluated L1 and L2 proficiency in speaking, reading, and writing. Jia at al. found that self-reported ratings of language proficiency were positively correlated with behavioral performance. Similarly, in a series of studies on how language dominance affects the degree of foreign accent and grammatical ability, Flege, MacKay, and Piske (2002) and Flege et al. (1999) used a language history questionnaire that targeted participants' self-reported age of arrival in the L2-speaking country/initial L2 learning; age of attained L2 proficiency; duration of L2 immersion; number of years of L2 schooling; percentage use of L1/L2; frequency of exposure to L2 TV, movies/videos, and radio; frequency of use of L1/L2 in a working environment; and ability to imitate foreign accents. Flege et al. (1999) found significant correlations between language history and degree of foreign accent in L2 and between language history and performance on a grammaticality judgment task (Flege et al., 2002). In a study of compuational language use, Vaid and Menon (2000) used a questionnaire that focused on language preference for mental arithmetic (e.g., counting, noting the time, remembering a telephone number). The questionnaire also yielded self-reported ratings of language proficiency speaking, comprehending, reading, and writing, as well as participants' age of arrival in the L2-speaking country/initial L2 learning; the setting of language acquisition (home, school, other); duration of L2 immersion; the language of instruction in elementary and secondary school; and frequency of use of L1/L2 at work, with parents, and with siblings, while thinking to self, and while dreaming. Language preference for mental arithmetic was found to correlate with variables in the bilinguals' language history, with the strongest predictor being the language of early formal instruction followed by length of residence in the L2 country, onset of bilingualism, and relative language dominance.

A consistent aspect of these studies was their focus on proficiency and history-related variables. However, the studies diverged in three notable ways: (a) The distinction among language proficiency, dominance, and preference remained largely unexplored, (b) behavioral tasks used to validate the questionnaires were limited, and (c) questions and scales were not consistent across studies. There is currently no uniform procedure for determining bilingual language dominance and proficiency. Researchers frequently use distinct aspects of language status and performance to delineate the two, or they use the same measure (e.g., L1:L2 proficiency ratio) to define both dominance (e.g., Flege et al., 2002) and proficiency (e.g., Vaid & Menon, 2000). For instance, whereas some studies have used self-ratings of ability to speak, understand, read, and write L1 and L2 (e.g., Goggin, Estrada, & Villarreal, 1994; Lemmon & Goggin, 1989; Magiste, 1979), others have relied on the experimenter's subjective judgment to determine dominance in a language (e.g., Talamas, Kroll, & Dufour, 1999). Factors related to bilinguals' language history, such as language of exposure in early years (e.g., Chincotta & Underwood, 1998; Hazan & Boulakia, 1993), current language use (e.g., Grosjean, 1982), bilinguals' speed of executing instructions (Lambert, 1955), and speed of naming pictures (Magiste, 1992), have all been used to operationalize language dominance. Whereas some researchers have assessed self-reported proficiency in comprehending, speaking, reading, and writing (e.g., Vaid & Menon, 2000), others have assessed only some, but not other, proficiency domains (e.g., Jia et al., 2002, who did not assess proficiency comprehending). Similarly, whereas some researchers have assigned bilinguals into dominance groups (L1 vs. L2) on the basis of vocabulary test scores in L1 and L2 (e.g., Cromdal, 1999), others have done so on the basis of speed of reading in the two languages (e.g., Favreau & Segalowitz, 1982; Macnamara, 1969). Adding to the confusion is that researchers have also, at times, used language preference, instead of language proficiency or dominance, as the domain of interest (e.g., Marian & Neisser, 2000; Ortiz & Garcia, 1990). Language preference is typically used to index participants' subjective feelings toward a language. In sum, some studies have focused on dominance ratings as indicators of true linguistic performance, but others have focused on proficiency or on preference ratings.

The second point of divergence among previous studies stems from ratings of language proficiency that were often compared to bilinguals' performance on only one or two behavioral tasks, such as degree of foreign accent (e.g., Flege et al., 1999) or grammaticality judgment (e.g., Jia et al., 2002), rather than on a range of behavioral tasks. There is considerable evidence that various language history variables apply differently across performance domains: For example, age of acquisition applies more to phonology but less to morphosyntax (Snow & Hoefnagel-Höhle, 1979). Therefore, an accurate picture of the relationship between bilinguals' self-reports and performance can be gained only from a comprehensive assessment of behavioral language performance.

Third, although some variables were assessed uniformly across studies (e.g., age of L2 learning), others were research specific (e.g., number of L1 and L2 speakers at home). Researchers target self-reported information that is relevant to the experimental manipulation at hand—for example, the ability to imitate a foreign accent is relevant to studying accents but is less relevant to studying computational language use. The different experimental goals often result in distinct questionnaires, making cross-experimental comparisons difficult. Finally, the scales used across studies often differ as well. For instance, whereas Jia et al. (2002) used a 4- or 5-point rating scale for all of their questions, Vaid and Menon (2000) elicited self-reported ratings on a 7-point scale.

Previous studies suggest that bilinguals' language profiles are best captured by assessing language experience and proficiency across multiple linguistic domains. It appears that bilinguals are able to assess their language proficiency and report their language history in a way that is consistent with behavioral performance (e.g., Chincotta & Underwood, 1998; Flege et al., 1999, 2002; Jia et al., 2002). However, the absence of a valid and uniformly used assessment measure makes it difficult to interpret existing findings and to make generalizations across studies and populations. In this article, we introduce a self-assessment tool that combines relevant proficiency and experience variables into a single instrument. We examine the relative value of each history variable for indexing bilinguals' actual linguistic performance across a range of standardized behavioral measures.

The Language Experience and Proficiency Questionnaire

The goal of this project was to develop a reliable and valid questionnaire for efficient assessment of bilinguals' linguistic profiles. Research indicates that ratings of proficiency alone are not sufficient to determine bilingual language status and that bilinguals' language learning and language use experiences play a significant role in shaping their linguistic competence (e.g., Grosjean, 2004; Hyltenstam & Abrahamsson, 2003). Therefore, the Language Experience and Proficiency Questionnaire (LEAP-Q) was constructed within the context of bilingualism theories that view L2 acquisition as an interplay between proficiency and experience variables (e.g., Hyltenstam & Abrahamsson, 2003). For example, Flege, Frieda, and Nozawa (1997) and Flege, Frieda, Walley, & Randazza (1998) demonstrated that bilinguals' dominance ratings were less predictive of L2 performance than experience-related variables and that although both language dominance and language history accounted for considerable variance in behavioral performance, language history variables were better predictors of L2 performance than dominance ratings (e.g., the frequency with which bilinguals spoke L1 influenced their performance in processing L2 phonology). It has been proposed that L2 experience variables become more important in shaping proficiency with increased L2 acquisition age (Hyltenstam & Abrahamsson, 2003; see also de Houwer, 1995) and that L2 acquisition is a result of cognitive, social, and environmental factors (e.g., Bialystok & Hakuta, 1994, 1999; Snow, 1983; Snow & Hoefnagel-Höhle, 1978). Therefore, in this study, both language proficiency and language history variables were considered necessary for specifying bilingual language status.

Given a theoretical framework that incorporates both language proficiency and language history, the LEAP-Q aims to capture factors that previously have been identified as important contributors to bilingual status: language competence (including proficiency, dominance, and preference ratings); age of language acquisition; modes of language acquisition; prior language exposure; and current language use. The LEAP-Q is based on question types previously used in questionnaires assessing bilinguals (e.g., Flege et al., 1999, 2002; Jia et al., 2002; Marian & Spivey, 2003; Vaid & Menon, 2000).

Language competence. Traditionally, the self-assessment literature has used three distinct measures to index bilingual language competence: (a) language proficiency, (b) language dominance, and (c) language preference. Because conflating the three measures can render interpretation of results difficult, each of them was probed separately in the LEAP-Q. Previous studies have construed proficiency as an index of general abilities across language processing domains (e.g., Bachman & Palmar, 1985; Stefani, 1994), including literacy-oriented proficiency, grammatical proficiency, vocabulary knowledge, and discourse abilities (Bachman, 1990; Harley, Cummins, Swain, & Allen, 1990). Consistent with other studies of bilingual self-assessment (e.g., Bahrick et al., 1994; Flege et al., 1999, 2002; Grosjean, 2004; Jia et al., 2002; Vaid & Menon, 2000), the LEAP-Q elicited proficiency ratings in speaking, listening, reading, and writing. However, instead of collapsing proficiency ratings along the different performance domains into a cumulative score (e.g., Flege et al., 2002), proficiency ratings obtained in this study were analyzed separately and were expected to yield different predictive information for different linguistic skills. For language dominance, participants in this study indicated dominance order for each of the languages spoken. The debate around the utility of a single global measure, such as language dominance (e.g., Oyama, 1978; Spolsky, Sigurd, Sako, Walker, & Arterburn, 1968), versus multiple task-specific measures, such as linguistic proficiency across domains (e.g., Bahrick et al., 1994; Fishman & Cooper, 1969), prompted the inclusion of both global (dominance) and specific (proficiency) measures of language competence, making it possible for the LEAP-Q to examine their effectiveness in indexing actual linguistic skills. Finally, LEAP-Q questions targeting preference were posed in specific terms (e.g., preference regarding reading a text available in all languages) rather than as general questions about the overall preferred language to maximize reliability and interpretation. Although a comparison of the three measures (proficiency, dominance, and preference) was not within the scope of this study, their availability in the LEAP-Q enables questionnaire users to weigh each of these measures against their variable of interest.

Language acquisition. Age of acquisition has been shown to be tightly connected to language learning, to influence bilinguals' ratings of language dominance, and to predict their performance on behavioral tasks (e.g., Hyltenstam & Abrahamsson, 2003; Johnson & Newport, 1989). For example, Flege et al. (2002) found that age of acquisition influenced bilinguals' dominance classification and correlated with bilinguals' sentence duration ratios in both languages. Consistent with studies demonstrating maturation effects in L2 acquisition, the LEAP-Q elicited four age-of-acquisition measures for each language spoken: (a) age of initial language learning, (b) age of attained fluency, (c) age of initial reading (i.e., age at which participants started to read in each language), and (d) age of attained reading fluency.

Moreover, the environment in which a language is learned also influences proficiency attainment. For instance, Flege et al. (1999) found that the number of years of education received in an L2 country, years of residence in an L2 country, average self-estimated use of L1 and L2, and chronological age all influenced age-of-acquisition effects on bilingual language dominance. The importance of environmental and contextual variables in language acquisition was demonstrated by Carroll (1967), who found a significant relationship between language performance and the extent to which the target language was used in the home. Therefore, the LEAP-Q elicits descriptions of acquisition modes in terms of the learning environments and in terms of the extent to which these learning environments contributed to language acquisition.

Prior and current language exposure. The degree of prior exposure to a language has been shown to influence research findings (e.g., Birdsong, 2005; Genesee, 1985; Kohnert, Bates, & Hernandez, 1999; MacKay & Flege, 2004; McDonald, 2000; Weber-Fox & Neville, 1999). For example, Flege et al. (1999) found that length of residence in the United States influenced bilinguals' sentence-level performance, with various language abilities differentially susceptible to language exposure (e.g., groups with more U.S. education had significantly higher rule-based morphosyntax scores than groups with less U.S. education but showed no differences in foreign accent ratings or in lexically based morphosyntax scores). Given the evidence that prior language exposure influences bilingual performance, the LEAP-Q assessed exposure to a language in four different environments: (a) in a country, (b) at school, (c) at work, and (d) at home.

In addition to prior exposure, ongoing language use can influence research findings. For example, Jia at al. (2002) found that mothers' L2 proficiency and frequency of speaking L2 at home were predictive of bilingual children's behavioral performance. Similarly, bilinguals who used L2 more often than L1 had better pronunciation and higher morphonosyntactic performance in L2 than bilinguals who used L1 more often than L2 (Flege et al., 2002). Therefore, the LEAP-Q elicited information regarding bilinguals' current exposure to their languages across settings, including interaction with family and friends and exposure during reading, watching TV, and listening to the radio, as well as exposure through self-instruction and language tapes.

LEAP-Q Aims

The target population for the LEAP-Q consists of adult and adolescent bilinguals and multilinguals with a variety of language experiences and proficiency levels. It includes simultaneous bilinguals (who learned their L1 and L2 early on and in parallel), late bilinguals (who learned their L2 later in life), balanced bilinguals (who are equally proficient across their languages), and unbalanced bilinguals (who are more proficient in one language than the other). In its current form, the LEAP-Q has been validated for use with individuals who have attained literacy skills equivalent to a high school education level or higher in at least one of their languages. Although previous studies suggest that similar questions about proficiency and language history can be successfully used to capture language profiles in bilingual children by means of parent reports (Chincotta & Underwood, 1998; Flege et al., 2002; Vaid & Menon, 2000), the questionnaire has not yet been validated for use with children or with clinical populations. Finally, unlike language testing instruments in which the primary objective is to place students in English-as-a-second language and foreign language programs (e.g., those offered by the Center for Advanced Research on Language Acquisition and the American Council on the Teaching of Foreign Languages, as well as the University of Ontario's French Immersion Program Assessment Tool), the target settings for the LEAP-Q are primarily research oriented (i.e., for assessment of research participants). Although novice language-learners may complete the questionnaire, the current study validated only self-reports of bilinguals with proficiency levels sufficient to complete standardized assessment tools (i.e., for self-reports to be meaningful, individuals should claim basic functionality within their L2). This target population was chosen for five reasons: (a) to be representative of bilingual and multilingual groups most often studied in research settings; (b) to accommodate the widest and most diverse sample of bilingual and multilingual speakers (with respect to specific languages, language history, and language use); (c) to allow the questionnaire to be completed independently and with minimal support while still providing meaningful data; (d) to span the variables documented as most relevant in surveys of practicing speech-language pathologists (e.g., Kohnert, Kennedy, Glaze, Kan, & Carney, 2003; Roseberry-McKibbin et al., 2005); and (e) to accommodate a trade-off between collecting a wide range of relevant measures while at the same time capturing sufficient detail about specific aspects of bilingual status.

The LEAP-Q was constructed to assess bilingual experience and proficiency profiles in first and second languages, irrespective of the specific languages involved. Therefore, in the current studies, different languages served as L1 and L2, with L1 status signifying that the language had been learned first (but not that it was dominant or that proficiency in it had been attained). Differences between subgroups of bilinguals (e.g., English–Spanish vs. Spanish–English) were not of interest in this research; instead, all data were analyzed in terms of L1 versus L2, thereby increasing the statistical power of various analyses and facilitating the questionnaire's generalizability to different populations of adult bilingual speakers. In both studies, the questionnaire was developed, administered, and normed in English only. Specific objectives of the project were to establish internal validity of the LEAP-Q, to establish criterion-based validity of the LEAP-Q, to establish predictive relationships between self-reported measures and performance on standardized language tests, and to establish predictive relationships between language history and self-reported proficiency levels.

Study 1: Establishing Internal Validity of the LEAP-Q

Study 1 aimed to establish internal validity of the questionnaire by analyzing responses of a diverse group of 52 bilinguals using factor analysis and multiple regression analyses. Factor analysis has been routinely used as a statistical method to uncover groups of variables that share variance patterns and are likely to measure the same construct. It has been successfully used in questionnaire and scale development, revealing constructs underlying cultural identification (e.g., Zea, Asner-Self, Birman, & Buki, 2003), emotional intelligence (e.g., Tapia, 2001), well-being of elderly bilinguals (e.g., Tran, 1994), and caring behaviors (e.g., Wu, Larrabee, & Putman, 2006). Factor analysis can uncover reliable underlying constructs despite possible surface dissimilarities in the data. Cross-linguistic studies have demonstrated that translating a questionnaire into a different language did not change the structure of factors, suggesting that the underlying constructs were not dependent on test language (e.g., Abdel-Khalek, Tomas-Sabado, & Gomez-Benito, 2004; Ferrer et al., 2006). For example, Wiebe and Penley (2005) showed that English scales translated into Spanish maintained the same factor structure. In Study 1, bilinguals' answers to LEAP-Q questions designed to measure a single construct were predicted to cluster together during factor analysis, to show distinct patterns for each language, and to yield factors indicative of L1 and L2 proficiency. It was also expected that bilinguals' language history would predict self-reported proficiency levels in L1 and L2.

Method

Participants. The questionnaire was administered to 52 multilingual individuals (M = 27.29 years, SD = 5.92; 29 women, 23 men). Participants were recruited from the Northwestern University campus and the greater Chicago metropolitan area communities. Participants varied in their education level from 2 years of college to a doctoral degree (M = 18.04 years of education, SE = 2.62; range = 15–27 years). None reported a hearing, language, or learning disability. Of the 52 participants, 11 spoke two languages, 19 spoke three languages, 12 spoke four languages, and 10 spoke five languages. Across participants, 34 languages were represented: American Sign Language, Belorussian, Bengali, Cantonese, Croatian, Czech, Dutch, English, Filipino, French, German, Hebrew, Hindi, Hungarian, Italian, Japanese, Korean, Mandarin, Malayalam, Marathi, Norwegian, Polish, Portuguese, Punjabi, Romanian, Russian, Spanish, Swahili, Tamil, Taiwanese, Telugu, Thai, Ukrainian, and Welsh.

Participants' self-reported language history and proficiency measures can be found in Table 1. L2 acquisition ages ranged from 0 to 15 years, representing both simultaneous and sequential bilinguals. Participants reported being exposed to L1 most in the context of family, followed by friends, reading, radio, TV, and independent language study. Participants reported being exposed to L2 most in the context of reading, followed by friends, TV, radio, family, and independent language study. When asked to report how different factors contributed to language learning, participants reported that learning L1 relied most on family, followed by friends, reading, TV, radio, and self-instruction, whereas learning L2 relied most on reading, followed by friends, TV, radio, family, and self-instruction (see Table 1). When asked to report proficiency in each language, participants reported highest proficiency for understanding, followed by speaking, reading, and writing for both L1 and L2.

Materials and procedure. The development of the questionnaire followed the steps outlined in *Questionnaires in Second Language Research* (Dörnyei, 2003, pp. 66–69) and included compilation of an initial item pool, discussion of questions, omission of jargon, and clarification and simplification of instructions and questions. First, an expert with experience in questionnaire development was consulted (the expert's primary area of expertise was audiology), and questionnaires in communication sciences and disorders–related fields were reviewed (e.g., Demorest & Erdman, 1987; Kohnert et al., 2003; Kuk, Tyler, Russel, & Jordan, 1990; Roseberry-McKibbin et al., 2005). Second, questions were chosen and prepared on the basis of assessment materials previously used in experimental studies conducted in our laboratory, other laboratories, and as described in the literature (Chincotta & Underwood, 1998; Flege et al., 1999, 2002; Jia et al., 2002; Vaid & Menon, 2000). Metacognitive judgment questions, such as evaluations of the relative contribution of different exposure variables to learning a language (e.g., interacting with family/friends, reading, etc.), also were included. Third, the draft version of the LEAP-Q was piloted with 8 bilingual and multilingual participants (Marian, Blumenfeld, & Kaushanskaya, 2003). All pilot participants spoke two or more languages and were representative of the LEAP-Q target population. Fourth, on the basis of participants' responses and feedback, the LEAP-Q was revised for clarity and succinctness to accommodate efficient self-reporting across different languages. Questions were considered on an item-by-item basis to ensure that none yielded missing values, outliers, or insufficient variability. Fifth, the resulting items were piloted with members of the Bilingualism and Psycholinguistic Research Laboratory and with members of the Communication Sciences and Disorders department at Northwestern University. On the basis of the feedback received, the LEAP-Q was revised and distributed to the participants in Study 1.

Domains assessed by the LEAP-Q included acquisition history, contexts of acquisition, present language use, language preference and proficiency ratings (across the four domains of language use: speaking, understanding, reading, and writing), and accent ratings. Some questions, such as those inquiring about ages of L2 acquisition, were applicable to all bilinguals tested; other questions, such as those inquiring about L1 learning from tapes, applied to a subgroup of bilinguals only (e.g., individuals in immigrant communities who learned their first language incompletely from their families and attempted to maintain and expand L1 proficiency by means of self-instruction). Questions pertaining to each language were designed to be identical, to accommodate variability in histories of L1 learning, and to maintain maximal flexibility of the questionnaire. Participants completed the questionnaire independently in approximately 25 min.

Analyses. Factor analysis was conducted to compare statistical clustering of questions with accepted dimensions of bilingual status. Seventy-seven attributes were entered into the principal components analysis, which served as the extraction method, and a varimax rotation method was applied. The statistical software was given a maximum of 100 iterations to converge on a factor solution, and the rotation converged in 51 iterations. Patterns of variables within a single construct were examined,

Table 1. Self-reported language history and proficiency for participants in Study 1.

Language history measures	L1 history			L2 history		
	M	*SD*	Range	*M*	*SD*	Range
Self-reported proficiency[a]						
Understanding	4.86	0.46	3.00–5.00	4.20	0.90	2.00–5.00
Speaking	4.65	0.60	3.00–5.00	4.02	0.99	1.00–5.00
Reading	4.25	1.52	0.00–5.00	4.20	1.11	0.00–5.00
Writing	4.08	1.48	0.00–5.00	3.76	1.39	0.00–5.00
Age milestones (years)						
Started learning	0.33	0.68	0.00–3.00	6.82	4.49	0.00–15.00
Attained fluency	3.61	2.59	0.00–10.00	14.31	8.90	0.00–38.00
Started reading	4.5	3.21	0.00–19.00	8.15	3.23	2.00–13.00
Attained reading fluency	8.07	2.94	3.00–15.00	15.10	7.74	5.00–43.00
Immersion duration (years)						
In a country	18.27	10.49	0.00–47.00	7.85	8.36	0.00–28.58
In a family	19.89	10.20	0.00–47.00	4.54	7.98	0.00–26.42
In a school	10.39	8.84	0.00–30.00	5.57	7.34	0.00–28.50
Contribution to language learning[b]						
From family	4.58	0.92	1.00–5.00	2.34	1.57	0.00–5.00
From friends	3.96	1.09	1.00–5.00	3.73	1.27	1.00–5.00
From reading	3.73	1.28	1.00–5.00	3.92	1.37	1.00–5.00
From TV	2.75	1.21	1.00–5.00	3.35	1.34	1.00–5.00
From radio	2.40	1.20	1.00–5.00	2.84	1.41	1.00–5.00
From self-instruction	1.43	1.04	0.00–5.00	1.92	1.21	0.00–5.00
Extent of language exposure[c]						
Family	4.32	1.04	1.00–5.00	2.00	1.30	1.00–5.00
Friends	3.59	1.30	1.00–5.00	3.35	1.28	1.00–5.00
Reading	3.00	1.47	1.00–5.00	3.67	1.53	1.00–5.00
TV	2.29	1.45	1.00–5.00	3.25	1.58	1.00–5.00
Radio	2.53	1.61	1.00–5.00	3.00	1.69	1.00–5.00
Independent study	1.08	0.35	1.00–3.00	1.24	0.66	1.00–4.00
Self-report of foreign accent[d]						
Perceived by self	0.56	1.10	0.00–5.00	1.87	1.49	0.00–5.00
Identified by others	0.38	0.79	0.00–4.00	2.19	1.42	0.00–4.00

Note. L1= first language; L2 = second language.

[a]Range: 0 (*none*) to 5 (*high*). [b]Range: 0 (*not a contributor*) to 5 (*most important contributor*). [c]Range: 1 (*not at all*) to 5 (*always*). [d]Range: 0 (*none*) to 5 (*very strong*).

commonalities underlying clusterings of variables within a single factor were identified, and the factor name was logically deduced. For factors that included variables with both positive and negative loadings, positive loadings provided inclusionary criteria and described the underlying construct reflected by the factor, whereas negative loadings provided exclusionary criteria and indicated an inverse relationship to the construct reflected by the factor. The names were intended to capture the nature of the variables that clustered together and to suggest underlying commonalities among them. This procedure is the standard way for performing factor analysis on large data sets (e.g., Tapia, 2001; Tran, 1994; Wu et al., 2006; Zea et al., 2003). In this study, although some constructs were expected to appear a priori, others were determined only after they emerged from analyses.

It should be noted that the factor analysis in this study was not used for combining items on the questionnaire into a single representative score. Although the LEAP-Q was constructed so that multiple questions measured a single domain of language proficiency (e.g., four questions targeted age of acquisition, four questions targeted proficiency), the questions were not intended to yield multi-item scales. For example, although the questions that clustered together in the L2 Competence factor were representative of a single underlying construct, combining these questions into a subscale of language competence may not be effective at measuring L2 proficiency in a different sample of bilingual participants. Theoretically, the construct of L2 Competence can be expected to involve at least three different subdomains: (a) age of acquisition, (b) length of immersion, and (c) self-reported degree of language proficiency. Differences across any of these domains would yield different bilingual profiles. Moreover, differences in population statistics would likely result in different factor structures across studies. Given the complex and varied nature of L2 acquisition across populations, questions in the LEAP-Q were intended to be considered separately and not to be reduced to a limited number of subscales. Factor analysis served as a tool for the determination of whether LEAP-Q questions contributed to underlying constructs shaping bilingual status and was, therefore, valid; it was not used to define these constructs as universal scales.

In addition, multiple regression analyses were conducted to examine the relationship between language history and language proficiency. Specifically, 16 attributes of language history (i.e., acquisition and fluency ages, learning environments, exposure variables) were entered as independent variables into stepwise multiple regression analyses, with self-reported proficiency in understanding, speaking, reading, and writing as dependent variables. Pearson R values, F and p values, and regression coefficients for the best predictor models are reported. (Note that magnitudes of beta values in regression analyses are, in part, dependent on the rating scale of the self-reported or behavioral measure in question.) Regression analyses in this study were used as exploratory tools rather than for theory testing. Although some researchers have found stepwise multiple regression analyses to be controversial (e.g., Menard, 1995; Tabachnick & Fidell, 2007), these analyses were the most appropriate way of examining the current data set because of the high number of independent variables (i.e., 16 language history variables). Stepwise multiple regression analyses made it possible to zero in on the variables that could account for the most variability in self-ratings, which is an especially useful approach when the number of independent variables is large (e.g., Mendenhall & Sincich, 1996). The predictive power of independent variables is marked by R values, where R values of .5 or higher are considered large, R values between .3 and .5 are considered moderate, and R values between .1 and .3 are considered small (Cohen, 1988).

Results

Sixteen factors with eigenvalues greater than 1 were extracted from the data set by means of factor analysis. Of these factors, the first 8 had eigenvalues greater than 3 and accounted for 76% of all variance (see Table 2). These 8 factors were assigned construct names indicative of their components and are listed in order of variance accounted for. Cronbach's alphas were calculated for each factor to assess consistency of components within each and to assess the extent to which questions captured similar information. (Cronbach's alphas should be .7 or higher for items in a set to be considered internally consistent with each other, although cutoffs ranging from .60 to .80 have been used; see Miller, 1995.)

The first factor, accounting for the most variance, included self-reported proficiency and comfort with speaking, understanding, reading, and writing in L1; identification with L1-associated culture; and preference for reading in L1 (all positive loadings), as well as L1 accent (negative loading; Cronbach's $\alpha = .85$). The positive loadings of L1 proficiency variables together with the negative loadings of L1 accent ratings suggested that this factor was an index of cross-modal *L1 Competence.*

The second factor (in order of variance accounted for) included age of initial L2 acquisition and age of attained L2 fluency (positive loadings), and comfort and proficiency in understanding L2 (negative loadings). Cronbach's alpha could not be calculated for this factor because of negative average covariance among items, which violated reliability model assumptions. The negative covariance resulted from the inverse relationship between age of acquisition and language competence—that is, later acquisition was associated with lower competence. Positive loadings for ages of acquisition suggested late language development, whereas negative loadings for proficiency variables suggested incomplete acquisition of the second language. Together, the clustering of these variables suggested that oral comprehension presented a special challenge for late learners, and this factor was interpreted as a measure of *Late L2 Learning.*

The third factor included total time exposed to L2; exposure to L2 via TV, friends, radio, family, reading, and classroom; proficiency and comfort with writing in L2; and a preference for speaking L2 (all positive loadings) as well as learning L2 from reading (negative loading; Cronbach's $\alpha = .92$). Positive loadings for L2 immersion and proficiency variables indicated a common underlying competence factor, whereas a negative loading for learning L2 from reading suggested that, in these bilinguals, competence in L2 was related to social L2 immersion. Together, these patterns were interpreted as an index of *L2 Competence.*

The fourth factor included total time exposed to L1; exposure to L1 via classroom, TV, radio, reading, and friends (positive loadings); and learning L2 from reading (negative loading; Cronbach's $\alpha = .80$). The presence of an L2 learning variable localized the underlying phenomenon to an L2 context, whereas the positive loadings of L1 exposure variables suggested continued immersion in L1. Therefore, this clustering was interpreted as a measure of continued L1 exposure—that is, *L1 Maintenance*.

Table 2. Factors yielded in Study 1.

Factor 1: L1 Competence	Loading values	Factor 2: Late L2 Learning	Loading values	Factor 3: L2 Competence	Loading values	Factor 4: L1 Maintenance	Loading values
Proficiency reading	.947	Age became fluent	.864	Exposure (% time)	.923	L1 exposure to classes	.916
Comfort understanding	.910	Age began acquiring	.859	Exposure to TV	.908	L1 exposure to TV	.914
Proficiency understanding	.910	Age became fluent reader	.855	Exposure to friends	.861	L1 exposure to radio	.831
Comfort writing	.903	Comfort understanding	−.803	Exposure to radio	.772	L1 exposure to reading	.776
Proficiency writing	.896	Age began reading	.751	Writing proficiency	.660	L2 learning from reading	−.727
Comfort reading	.884	Proficiency understanding	−.697	Exposure to family	.621	L1 exposure (% time)	.627
Identified accent	−.788	Years in a country	−.681	Comfort writing	.592	L1 exposure to friends	.530
Comfort speaking	.748	Learning from tapes	.601	Preference to speak	.590		
Proficiency speaking	.704	Proficiency speaking	−.580	Exposure to reading	.564		
Cultural identification	.526			Exposure to classes	.543		
Perceived accent	−.517			Learning from reading	−.519		
Preference to read	.457						
% variance	23.480		13.383		9.625		7.534
Cumulative variance	23.480		36.862		46.488		54.021

Factor 5: Late L2 Immersion	Loading values	Factor 6: Media-Based Learning	Loading values	Factor 7: Non-Native Status	Loading values	Factor 8: Balanced Immersion	Loading values
L1 years of class learning	.728	L1 learning from TV	.866	L2 perceived accent	.839	L1 learning from friends	−.813
L2 years in workplace	.725	L2 learning from TV	.838	L2 identified accent	.615	L2 years of schooling	.627
L1 years in workplace	.714	L1 learning from the radio	.741	L2 cultural identification	−.602	L1 years in a family	.622
Proficiency reading L2	−.687	L2 learning from the radio	.652	L1 age became fluent	.590	L2 years in a classroom	.541
L2 learning from friends	−.683	L2 comfort reading	−.476	L2 learning from family	−.519	L1 years in a country	.499
L2 learning in a classroom	−.556						
L1 years in school	.476						
% variance	6.424		6.226		5.049		4.232
Cumulative variance	60.445		66.671		71.720		75.952

The fifth factor included years exposed to L1 in a language classroom, workplace, and general school setting, and years exposed to L2 in the workplace (all positive loadings), as well as proficiency in reading L2 and learning L2 from friends and in a foreign language classroom (negative loadings; Cronbach's α = .30; however, note that Cronbach's alphas are likely to be underestimated in factors where both positive and negative loadings are present). The positive loading of L2 workplace immersion suggested adult L2 acquisition, whereas positive loadings of a range of L1 immersion variables (including workplace and school settings) suggested late immigration from the L1-speaking country. This clustering of variables is likely to describe a subset of bilinguals consisting of adult immigrants, who spent their formative years in an L1-speaking environment and who were immersed in L2 later in life. To account for the presence of both L1 and L2 variables within this cluster, this factor was interpreted to index *Late L2 Immersion*.

The sixth factor included learning L1 and L2 from radio and TV (positive loadings), as well as comfort with reading in L2 (negative loading; Cronbach's α = .75). Positive loadings for both L1 and L2 suggested that the construct underlying this factor was not specific to just one language; instead, it was more likely to capture a general trend of language learning from the media. This pattern of variable loadings indicated a measure of language learning within a popular culture framework and was interpreted as a measure of *Media-Based Learning*.

The seventh factor included L2 accent as perceived by the participant and as identified by others, age of attained L1 fluency (positive loadings), and identification with L2 culture and learning L2 from family (negative loadings; Cronbach's α = .24). Positive loadings for acquisition ages and for accent judgments indicated late L2 acquisition, and negative loadings for identification with L2 culture and learning L2 from family indicated lack of assimilation into the L2 environment. Age of attained L1 fluency loaded positively, suggesting that this cluster was specific to a subgroup of bilinguals who experienced a discontinuity in use of their L1. Together, these patterns were interpreted to index *Non-Native Status*.

The eighth factor included years spent in an L1 family and country, years exposed to L2 in a language classroom and in a general school setting (all positive loadings), and learning L1 from friends (negative loading; Cronbach's α = .27). Positive loadings from both L1 and L2 variables suggested that the underlying phenomenon was common to both languages, whereas the negative loading for learning L1 from friends suggested relatively early immigration from the L1-speaking country. Together, these patterns suggested a balanced bilingual profile, where both L1 immersion and L2 immersion were important. Therefore, this factor was probably descriptive of a subset of bilinguals who were born in an L1 country but were schooled mostly in an L2 country, and it was interpreted to reflect *Balanced Immersion*.

The language history measures that predicted proficiency in understanding, speaking, reading, and writing in L1 and L2 are reported in Table 3. The results of the stepwise multiple regression analyses include regression coefficients (marking the relative importance of each independent variable that entered the model) and statistics describing the fit of the model. In addition to standardized regression beta, variation inflation factors (VIFs) are included to measure the correlation among independent variables (lack of such correlation is a basic assumption in regression analysis). It is generally accepted that VIF values greater than 10 signal multicollinearity and singularity problems (Mendenhall & Sincich, 1996). In this study, VIF values of independent variables ranged from 1.0 to 1.4, suggesting that no multicollinearity/singularity problems were present.

Discussion

Participants in Study 1 reported high levels of proficiency and extensive immersion in both languages. Across modalities, L1 proficiency was higher than L2 proficiency, with the largest difference in understanding and the smallest difference in reading. It is possible that the emphasis on reading in this study is reflective of the specific sample of bilinguals in Study 1, who were recruited primarily from academic communities. In order to succeed in an L2 academic environment, reading in L2 was likely to be particularly important. Therefore, bilinguals in this study may have been more likely to judge their overall L2 competence on the basis of L2 reading skills versus other skills. The emphasis on L2 reading evident in proficiency ratings was echoed in participants' reports that reading experiences (e.g., learning via reading and exposure to reading) contributed most to L2 competence, whereas family-based experiences (e.g., learning language from family and exposure to family) contributed most to L1 competence. These patterns of reading-oriented L2 acquisition accurately reflect typical language learning patterns for sequential bilinguals, where a native language is acquired within a family environment and a second language is often acquired on entrance into schooling and takes place in a classroom environment involving explicit reading instructions.

Factor analysis yielded component groupings that accounted for most of the variance in bilinguals' self-reported data, suggesting that questions on the LEAP-Q were broad enough to capture variability in the bilingual population that was sampled in this study. The clusters that emerged reflected underlying dimensions of bilingualism, with questions that clustered together measuring the same construct. The first four factors (L1 Competence,

Table 3. Multiple regression analyses for Study 1: Language history predictors of proficiency, including regression coefficients (B and β) and corresponding variance inflation factors (VIFs), as well as evaluators of the fit of the model (R, R^2, and Fs).

Predictee, *F* test	Predictor	Regression coefficients				Fit of model	
		B	*SE* of B	β	VIF	*R*	R^2
Understanding L1	Years in an L1 country	0.02	0.01	.44	1.09	.55	.31
$F(2, 50) = 12.7$, $p < .001$	Learning L1 from reading	0.14	0.05	.40	1.09	.67	.45
Speaking L1	Years in an L1 country	0.04	0.01	.55	1.04	.61	.38
$F(2, 50) = 15.1$, $p < .001$	Exposure to L1 friends	0.17	0.06	.35	1.04	.70	.49
Reading L1	Years in an L1 country	0.06	0.02	.41	1.09	.52	.27
$F(2, 50) = 11.5$, $p < .001$	Learning L1 from reading	0.45	0.16	.41	1.09	.65	.43
Writing L1	Years in an L1 country	0.11	0.02	.61	1.09	.74	.55
$F(2, 50) = 29.4$, $p < .001$	Learning L1 from reading	0.46	0.13	.40	1.09	.83	.69
Understanding L2	Age began reading in L2	−0.12	0.03	−.52	1.00	.52	.27
$F(1, 50) = 11.6$, $p < .01$							
Speaking L2	L2 exposure in family	0.35	0.09	.57	1.33	.48	.23
$F(4, 50) = 7.4$, $p < .001$	Age began reading in L2	−0.12	0.04	−.45	1.20	.60	.35
	L2 classroom learning	0.28	0.10	.42	1.32	.66	.43
	L2 exposure in classroom	−0.14	0.07	−.32	1.43	.71	.51
Reading L2	Learning L2 from reading	0.35	0.10	.50	1.07	.51	.26
$F(3, 50) = 9.8$, $p < .001$	Learning L2 from friends	0.34	0.11	.44	1.13	.60	.36
	Learning L2 from radio	−0.24	0.08	−.39	1.13	.70	.50
Writing L2	Exposure to L2 friends	0.52	0.11	.64	1.05	.58	.33
$F(2, 50) = 11.5$, $p < .001$	Years in an L2 workplace	−0.06	0.03	−.31	1.05	.65	.43

Late L2 Learning, L2 Competence, and L1 Maintenance) were general in nature, provided information shared by all bilinguals, and accounted for more than half of the variance in the data. The remaining factors (Late L2 Immersion, Media-Based Learning, Non-Native Status, and Balanced Immersion) grew increasingly specific and appeared to be driven by bilingual subgroups. In addition, the measures of variance in the data set accounted for by each factor (indicated by eigenvalues) were augmented by measures of consistency within each observed factor (indicated by Cronbach's alphas). The four factors with highest Cronbach's alphas were general in nature: L1 Competence, L2 Competence, L1 Maintenance, and Media-Based Learning. Conversely, the three factors with lower Cronbach's alphas—Late L2 Immersion, Non-Native Status, and Balanced Immersion—might have emerged from characteristics of specific subgroups in the sample tested. Together, the eigenvalues and Cronbach's alphas reveal that L1 Competence, L2 Competence, and L1 Maintenance accounted for much of variance in the data and were highly consistent internally. (Late L2 Learning accounted for approximately 13% of the variance but did not lend itself to statistical evaluation of internal consistency because of negative covariance across age and ability components. Finally, Media-Based Learning had high internal consistency but accounted for relatively little variance in the data, about 6%.)

The structure of the L1 Competence and L2 Competence factors suggests that the constructs of proficiency in the first and second language share subcomponents. However, some differences in factorial structures were also observed. For instance, unlike the L1 Competence factor, the L2 Competence factor included a large number of age-of-acquisition variables. The significant contribution of maturation variables to competence in L2 is consistent with previous studies that have identified L2 acquisition age as an important determiner of L2 competence (e.g., Flege et al., 1999, 2002; Jia et al., 2002; Johnson & Newport, 1989; Vaid & Menon, 2000). Moreover, the structure of the L2 Competence factor confirms Hyltenstam and Abrahamsson's (2003) model of L2 acquisition, which postulates interplay between age of acquisition and environmental variables in shaping L2 attainment.

Although some factors may include variables that appear similar, they represent constructs that differ in nuanced ways. For example, L2 learning in a classroom loads negatively onto Factor 5 (Late L2 Immersion), and years in an L2 classroom loads positively onto Factor 8 (Balanced Immersion). The two questions may contribute to overall language proficiency in distinct ways. Classroom experience as a contributor to learning L2 is a subjective, metacognitive self-assessment measure, whereas duration of L2 classroom exposure is an objective temporal measure. Contribution to learning may reflect learning style, whereas years of exposure measures duration of experience (amount of time spent in a classroom is not always indicative of language proficiency). This distinction is reflected in differential loadings of these two variables onto Factors 5 and 8.

Multiple regression analyses were used to generate predictive equations for self-reported proficiency levels in L1 and L2. The results suggested that different experiential variables predicted proficiency in the two languages. For instance, L1 proficiency levels across modalities were consistently predicted by years spent in an L1 country, whereas L2 proficiency levels were predicted by L2 acquisition ages. In contrast to L1 regression models, years of immersion (in this case, length of time spent in an L2 workplace) predicted only proficiency in writing L2, but not proficiency in speaking, understanding, or reading L2. Moreover, although similar experiential variables predicted proficiency levels in L1 across understanding, reading, speaking, and writing, predictors of proficiency levels in L2 were more varied. For instance, L1 proficiency levels were consistently predicted by time spent in an L1 country. Conversely, predictors of L2 proficiency differed: For example, proficiency in understanding was predicted primarily by age when participants began reading L2, and proficiency in speaking was predicted primarily by exposure to L2 in family and classroom environments. The greater variability and larger number of language history predictors for L2 is likely due to different acquisition patterns for the two languages, with L2 acquisition more varied across settings and modalities relative to L1. Greater variability in L2 acquisition patterns is probably due to highly diverse experiences associated with learning a second language (relative to a native language, the acquisition of which is less variable). For example, whereas some participants learned L2 in an immersion-type setting (because of immigration, or studying and working abroad), others learned L2 in a classroom environment.

The results of Study 1 prompted exclusion of four measures from the questionnaire. First, measures of comfort across modalities (e.g., "How comfortable are you speaking, understanding, reading, and writing in a language?") were excluded. These questions yielded values that were similar to those yielded by proficiency measures—that is, the two patterned identically, correlated significantly, and were highly predictive of each other. Specifically, proficiency measures were predictive of comfort measures for speaking (L1: R = .9, F[1, 50] = 167, p < .001; L2: R = .9, F[1, 50] = 264, p < .001), understanding (L1: R = .95, F[1, 50] = 468, p < .001; L2: R = .9, F[1, 50] = 332, p < .001), reading (L1: R = .97, F[1, 50] = 881, p < .001; L2: R = .9, F[1, 50] = 181, p < .001), and writing (L1: R = .99, F[1, 50] = 1760, p < .001; L2: R = .96, F[1, 50] = 561, p < .001). Second, writing proficiency was excluded because of its close relationship to reading proficiency—specifically, proficiency in reading L1 predicted proficiency in writing L1, R = .9, F(1, 50) = 365, p < .001, and proficiency in reading L2 predicted proficiency in writing L2, R = .8, F(1, 50) = 108, p < .001. The high predictability of writing proficiency from measures of reading proficiency suggested that a separate question assessing writing proficiency would not provide additional new information. Third, current classroom exposure was excluded because of its close relationship to current reading exposure, with reading exposure in L1 predicting ratings of classroom exposure in L1, R = .7, F(1, 50) = 34, p < .001, and reading exposure in L2 predicting ratings of classroom exposure in L2, R = .6, F(1, 50) = 33, p < .001. Because classroom exposure was generally not predictive of self-reported proficiency ratings, and because the extent of classroom exposure could be reliably deduced on the basis of reading exposure, it was omitted from the final version of the LEAP-Q. Fourth, percentage of bilingual contacts was excluded because it did not correlate with any other measures and did not load onto any factor in the factor analysis. Because the aim of this study was to construct an internally consistent self-assessment measure, a question that was not related to any other questions in the questionnaire was deemed unreliable and uninformative. As a result of these changes, seven questions per language were omitted, shortening questionnaire completion time without losing predictive value. In addition, because of limited variability (1–5) in questions that required responses on a scale, the ranges of values on all scales were increased (1–10). Finally, the questionnaire was transferred into digital format. Pull-down menus were added to questions that required responses on scales, thus facilitating questionnaire completion and subsequent data extraction for analyses. In the digitized version, once the participant indicated a specific language, the language was filled in automatically throughout the questionnaire. The revised version of the questionnaire was used in Study 2 and required approximately 15 min for bilinguals to complete.

Study 2: Establishing Criterion-Based Validity

The objectives of Study 2 were to confirm the internal validity of the LEAP-Q in a more homogeneous sample of bilinguals through factor analysis as well as to establish criterion-referenced validity by comparing self-reported and standardized proficiency measures using correlation and regression analyses. We made the following five predictions:

1. Bilinguals' answers to questions that referred to the same underlying aspects of bilingual profiles would

pattern together in factor analysis and would mirror factors revealed in Study 1, therefore supporting the internal validity of the questionnaire.

2. Similar to Study 1, bilinguals' self-reported language history would be predictive of their self-reported proficiency levels in L1 and L2.
3. Self-reported proficiency would correlate with and predict performance on standardized language measures.
4. In turn, performance on standardized language tests would predict self-reported proficiency, providing information about the metaknowledge on which bilinguals rely when estimating their language proficiency.
5. Language history would predict performance on standardized language measures.

Method

Participants. Fifty bilingual speakers of English and Spanish participated (M = 26.7, SD = 10.4; 31 women, 19 men). Of these, 18 were Spanish–English bilinguals (native speakers of Spanish who acquired English as a second language), and 32 were English–Spanish bilinguals (native speakers of English who acquired Spanish as a second language). Participants were recruited from the Northwestern University and the Chicago metropolitan area communities by means of flyers. None reported a hearing, language, or learning disability. L1 and L2 acquisition ages, language histories, and language performance information are presented in Tables 4 and 5. L2 acquisition ages ranged from 0 to 23 years of age,

Table 4. Self-reported language history and proficiency for participants in Study 2.

Language history measures	L1 history			L2 history		
	M	*SD*	Range	*M*	*SD*	Range
Self-reported proficiency[a]						
Understanding	9.58	0.93	6.00–10.00	7.92	2.33	2.00–10.00
Speaking	9.32	1.15	6.00–10.00	7.74	2.05	2.00–10.00
Reading	9.26	1.26	5.00–10.00	8.02	1.97	2.00–10.00
Age milestones (years)						
Started learning	1.08	1.75	0.00–11.00	8.25	5.95	0.00–23.00
Attained fluency	4.00	2.81	0.00–13.00	15.10	7.44	2.00–39.00
Started reading	5.5	2.24	3.00–14.00	11.52	5.44	3.00–25.00
Became fluent reading	7.92	3.27	4.00–41.00	15.09	7.44	2.00–39.00
Immersion duration (years)						
Country	17.6	10.5	0.00–54.00	11.3	12.2	0.00–44.00
Family	22.65	7.96	0.00–55.00	8.72	12.18	0.00–37.00
School	17.62	8.46	3.00–49.00	9.55	10.66	0.00–49.00
Contribution to language learning[b]						
From family	8.94	1.97	1.00–10.00	3.58	3.76	0.00–10.00
From friends	7.3	2.74	0.00–10.00	6.26	3.12	0.00–10.00
From reading	7.54	2.69	0.00–10.00	7.32	2.03	3.00–10.00
From TV	4.88	2.85	0.00–10.00	5.08	3.11	0.00–10.00
From radio	3.14	2.84	0.00–10.00	3.64	3.39	0.00–10.00
From self-instruction	1.16	2.58	0.00–10.00	3.32	3.63	0.00–10.00
Extent of language exposure[c]						
To family	7.62	3.02	0.00–10.00	3.16	3.46	0.00–10.00
To friends	7.34	3.34	0.00–10.00	5.18	3.81	0.00–10.00
To reading	6.26	3.33	0.00–10.00	5.56	3.43	0.00–10.00
To TV	6.16	3.47	0.00–10.00	4.6	3.87	0.00–10.00
To radio	6.48	3.23	0.00–10.00	4.9	3.70	0.00–10.00
Self-instruction	1.32	2.89	0.00–10.00	2.18	3.02	0.00–10.00
Self-reported foreign accent[d]						
Perceived by self	0.70	1.25	0.00–6.00	3.28	2.62	0.00–9.00
Identified by others	0.82	1.85	0.00–8.00	4.06	3.38	0.00–10.00

[a]Range: 0 (*none*) to 10 (*perfect*). [b]Range: 0 (*not a contributor*) to 10 (*most important contributor*). [c]Range: 0 (*never*) to 10 (*always*). [d]Range: 0 (*none*) to 10 (*pervasive*).

Table 5. Standardized proficiency measures for Study 2.

Measure	L1 performance			L2 performance			L1–L2 comparisons
	M	*SD*	Performance range	*M*	*SD*	Performance range	*t* tests and effect sizes (partial η^2)
Woodcock–Johnson/Woodcock–Muñoz							
Reading Fluency (percentile)	62	28	12–99.9	35	27	2–97	$t(49) = 4.56$, $p < .01$, $\eta^2 = .30$
Oral Comprehension (percentile)	64	21	3–98	36	30	0.1–97	$t(49) = 4.78$, $p < .01$, $\eta^2 = .33$
Passage Comprehension (percentile)	59	26	5–96	38	31	0.1–93	$t(49) = 3.32$, $p < .01$, $\eta^2 = .18$
Productive Vocabulary (percentile)	47	28	0.1–99.9	20	23	0.1–75	$t(49) = 4.51$, $p < .01$, $\eta^2 = .29$
Sound Awareness (percentile)	50	20	19–95	35	22	6–89	$t(49) = 3.74$, $p < .01$, $\eta^2 = .26$
PPVT/TVIP measures (percentile)							$t(49) = 4.76$, $p < .01$, $\eta^2 = .32$
	80	16	42–99.9	60	27	0.1–96	
Grammaticality judgments							
Accuracies (percentage)	80	20	38–100	70	20	31–96	$t(49) = 3.26$, $p < .01$, $\eta^2 = .18$
Response latencies (ms)	2,786	1,033	980–5,290	3,842	1,445	1,781–9,861	$t(49) = 4.01$, $p < .01$, $\eta^2 = .25$

Note. PPVT = Peabody Picture Vocabulary Test; TVIP = Test de Vocabulario en Imágenes Peabody.

representing both simultaneous and sequential bilinguals. Participants varied in their education levels from high school to graduate school (M = 16 years of education, SE = 2.5, range = 11–22 years; note that in some countries, a high school education is equivalent to fewer than 12 years). Participants reported being exposed to L1 most in the context of family, followed by friends, radio, reading, TV, and self-instruction. They reported being exposed to L2 most in the context of reading, followed by friends, radio, TV, family, and self-instruction. When asked to report how different factors contributed to language learning, participants reported that learning L1 relied most on family, followed by reading, friends, TV, radio, and self-instruction, and they reported that learning L2 relied most on reading, followed by friends, TV, radio, family, and self-instruction. Participants performed better in L1 than in L2 across a range of behavioral measures (see Table 5 for means and statistical comparisons).

Materials and procedure. The revised version of the LEAP-Q was administered to all participants at the start of the experimental session on a computer. Participants independently completed the LEAP-Q in English. On completion, participants were administered a battery of standardized behavioral measures of language ability. These included subtests from the Woodcock–Johnson Tests of Achievement (Woodcock, McGrew, & Mather, 2001), the Woodcock–Muñoz Tests of Achievement (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005), and the Peabody Picture Vocabulary Test in English (PPVT, Dunn & Dunn, 1997) and Spanish (Test de Vocabulario en Imágenes Peabody [TVIP]; Dunn, Padilla, Lugo, & Dunn, 1986). In addition, sentence grammaticality judgment tasks were constructed on the basis of previous materials (Bedoya et al., 2005; DeKeyser, 2000; Johnson & Newport, 1989). All measures were administered in language blocks, with half the participants receiving the Spanish measures first and half receiving the English measures first. Test administrators were highly proficient in both languages. Specifically, the following seven behavioral measures were administered:

1. A reading fluency test (Subtest 2 of the Woodcock–Johnson Test of Achievement in English and its equivalent Woodcock–Muñoz version in Spanish). This test required participants to read as many sentences as possible within a 3-min interval and to decide whether each sentence was true or false.
2. A passage comprehension test (Subtest 9 of the Woodcock–Johnson Test of Achievement in English and its equivalent Woodcock–Muñoz version in Spanish). This test required participants to read passages and supply missing words.
3. A productive picture vocabulary test (Subtest 14 of the Woodcock–Johnson Test of Achievement in English and its equivalent Woodcock–Muñoz version in Spanish). This test required participants to name pictures.
4. An oral comprehension test (Subtest 15 of the Woodcock–Johnson Test of Achievement in English and its equivalent Woodcock–Muñoz version in Spanish). This test required participants to listen to passages and supply missing words.
5. A sound awareness test (Subtest 21 of the Woodcock–Johnson Test of Achievement in English and its equivalent Woodcock–Muñoz version in Spanish). This test required participants to complete a rhyming task, a sound deletion task, a sound substitution task, and a sound reversal task.

6. A receptive vocabulary test (PPVT/TVIP). This test required participants to identify pictures in response to auditory instructions.
7. A grammaticality judgment test (based on Bedoya et al., 2005; DeKeyser, 2000; Johnson & Newport, 1989). This test required participants to read 50 sentences in English and 50 sentences in Spanish and to judge whether these sentences were grammatically correct. For each sentence, a grammatically correct and a grammatically incorrect version were presented at least 20 trials apart from each other. The test was administered on a computer screen, and participants identified sentences as correct or incorrect by pressing keys on the keyboard. Sentences were matched in length, both in terms of number of letters (because the test was administered visually) and in terms of number of phonemes (in case subvocal articulation strategies were used). The number of letters in English (M = 30.0, SD = 7.3) and Spanish sentences (M = 28.7, SD = 7.3) did not differ, $t(98) = 0.9$, $p > .1$; the number of phonemes in English (M = 24.7, SD = 5.7) and Spanish sentences (M = 27.0, SD = 7.0) also did not differ, $t(98) = 1.8$, $p > .05$.

Data Coding and Analyses

For all standardized measures, participant scores were coded in terms of percentile rankings on a scale from 0 to 100. Percentile rankings were derived following the normative data available for each of the standardized tests (PPVT/TVIP [see Dunn & Dunn, 1997, and Dunn et al., 1986] and the Woodcock–Johnson Tests of Achievement and Woodcock–Muñoz Tests of Achievement [see Woodcock et al., 2001, and Muñoz-Sandoval et al., 2005]). Grammaticality judgment accuracy was coded as proportion correct, and grammaticality judgment latencies were coded in milliseconds. Grammaticality judgment latencies were measured as the duration of time between the onset of sentence presentation and the participant's response. Only the latencies for correctly identified sentences were included in the latency analyses. Outliers (latencies longer or shorter than 3 SDs from the participant's mean) were excluded from analyses.

Factor analysis, as well as correlation and multiple regression analyses, were conducted. Multiple regression analyses were run to examine predictors of proficiency levels in bilinguals. First, 16 attributes of language history were entered as independent variables into stepwise multiple regression analyses, and self-reported proficiency in understanding, speaking, and reading were entered as dependent variables. Second, questionnaire-based language history attributes were entered as independent variables into stepwise multiple regression analyses, and behaviorally established proficiency measures were entered as dependent variables. Third, results of standardized language tests were entered as independent variables into stepwise multiple regression analyses, and self-reported proficiencies in understanding, speaking, and reading were entered as dependent variables.

Results

Fourteen factors with eigenvalues greater than 1 were extracted from the data set using factor analysis. Of these factors, the first 5 had eigenvalues greater than 3. For consistency across the two studies, the first 8 factors were assigned construct names indicative of their components and are listed in order of variance accounted for. Each of these factors had eigenvalues greater than 2.2, and together they accounted for 73.5% of variance in the data (see Table 6). Cronbach's alpha was calculated for each factor separately, yielding values ranging from .31 to .92, suggesting overall consistency of components within each factor.

The first factor (accounting for the most variance) included total time exposed to L2; exposure to L2 via TV, reading, friends, and radio; years spent in an L2 country; proficiency reading and speaking L2; and preference to read in L2 (all positive loadings), as well as total time exposed to L1; exposure to L1 via reading, TV, radio, and friends; preference to read in L1; learning L1 from reading; and accent in L2 as identified by others (all negative loadings). The observed distinct patterns for L2 versus L1 suggested separate subcomponents (inversely related) for each language within a single factor. The finding that in this factor, L2-related variables loaded positively and L1-related variables loaded negatively may indicate that this group of bilinguals was L2 dominant. Moreover, the inverse relationship between L1 and L2 variables is consistent with other studies of relative L1/L2 language competence (e.g., Flege et al., 2002; Harley et al., 1990). To circumvent a violation of assumptions for obtaining reliability values (i.e., negative covariance between L2 and L1 components), Cronbach's alpha was computed separately for L2 (.88) and L1 (.92) components. This factor was taken to index *Relative L2–L1 Competence*.

The second factor (in order of variance accounted for) included age that participant began learning to read in L1; age of becoming a fluent L1 reader; and learning L1 from tapes (all positive loadings); as well as proficiency in speaking, reading, and understanding L1 (negative loadings; Cronbach's α = .31). Positive loadings of acquisition age variables suggested a late-acquisition profile, whereas negative loadings of proficiency variables suggested an incompletely acquired L1. This L1 learning profile may be characteristic of bilinguals who immigrated from an L1-speaking country early in life and/or of bilinguals who

Table 6. Factors yielded in Study 2.

Factor 1: Relative *L2–L1 Competence*	Loading values	Factor 2: L1 Learning	Loading values	Factor 3: Late L2 Learning	Loading values	Factor 4: L1 Nondominant Status	Loading values
L1 exposure to reading	−.872	Speaking proficiency	−.929	L2 learning from the radio	.755	L1 age began acquiring	.871
L2 exposure to TV	.866	Age began reading	.824	L2 age when fluent reader	.724	L1 learned from family	−.823
L2 exposure to reading	.844	Reading proficiency	−.747	L2 age when fluent	.654	L1 age when fluent	.624
L1 exposure to TV	−.811	Age when fluent reader	.745	L2 age began reading	.605	L2 preference to speak	.577
L2 exposure to friends	.804	Comprehension proficiency	−.709	L2 exposure to self-instruction	.561	L1 identified accent	.526
L2 exposure to radio	.777	Learning from tapes	.629	L2 learning from tapes	.557	L1 exposure to family	−.468
L1 exposure to radio	−.773			L1 perceived accent	−.548		
L1 exposure (% time)	−.771			L2 perceived accent	.452		
L1 exposure to friends	−.762						
L2 exposure (% time)	.759						
L2 reading proficiency	.754						
L1 preference to read	−.713						
L2 preference to read	.696						
L2 speaking proficiency	.580						
L1 learning from reading	−.522						
L2 identified accent	−.496						
L2 years in the country	.474						
% variance	25.296		12.425		9.615		7.471
Cumulative variance	25.296		37.722		47.337		54.807

Factor 5: L2 Immersion	Loading values	Factor 6: L1 Immersion	Loading values	Factor 7: L2 Nonacculturation	Loading values	Factor 8: Media-Based L1 Learning	Loading values
Exposure to family	.898	L1 years of schooling	.885	L2 age when began acquiring	0.854	L1 Learning from radio	.845
Years in family	.894	L1 years in family	.818	L2 cultural identification	−0.713	L1 Learning from TV	.816
Learning from family	.747	Chronological age	.766				
Years of schooling	.537	L1 years in country	.740				
		L2 learning from TV	−.503				
% variance	5.933		4.860		4.095		3.835
Cumulative variance	60.740		65.600		69.695		73.530

learned L1 at home but lived and received schooling from an early age in an L2 environment and made an active effort to maintain L1. These patterns were interpreted as reflecting *L1 Learning*.

The third factor included learning L2 from radio and language tapes, exposure to L2 through independent study, ages of becoming a fluent L2 speaker and reader, and self-perceived accent in L2 (all positive loadings), as well as self-perceived accent in L1 (negative loading; Cronbach's α = .77). Positive loadings of acquisition age and self-instruction variables suggest late and incomplete acquisition of the second language, whereas positive loadings of L2 accent and negative loadings of L1 accent suggest higher L1 fluency than L2 fluency. Together, these patterns were interpreted as indexing *Late L2 Learning*.

The fourth factor included ages of L1 acquisition, ages of attained L1 fluency, L1 accent as identified by others, and preference to speak L2 (all positive loadings), as well as exposure to and learning from an L1-speaking family (negative loadings), suggesting limited L1 exposure. Cronbach's alpha could not be calculated for this factor because of the negative average covariance among items, a violation of reliability model assumptions. Positive loadings of acquisition age and accent variables indicate lack of fluency in L1, whereas negative loadings of family exposure in L1 suggest lack of immersion in an L1 environment. These variables are likely to describe a subset of bilinguals for whom L1 is no longer a dominant language and who prefer to use L2 in daily life. Therefore, this factor was interpreted as indexing *L1 Nondominant Status*.

The fifth factor included L2 family-based components (e.g., current exposure to, years spent in, and learning from an L2-speaking family) as well as years of L2 schooling (all positive loadings; Cronbach's α = .86). This factor was interpreted to suggest an overall measure of interactive *L2 Immersion*.

The sixth factor included years spent in an L1-speaking school, family, and country; participant's chronological age (all positive loadings); and learning L2 from TV (negative loading: Cronbach's α = .50). The negative loading of L2 learning variables may be indicative of a subset of bilinguals who maintained their first language and had minimal exposure to their second language. This factor was interpreted as an overall measure of interactive *L1 Immersion*.

The seventh factor included age of L2 acquisition (positive loading) as well as identification with L2 culture (negative loading). Cronbach's alpha could not be calculated for this factor because of the negative average covariance among items, a violation of reliability model assumptions. The positive loading of L2 acquisition age onto this factor indicated a late L2 learning profile, whereas the negative loading of L2 cultural identification suggested lack of acculturation within the L2-speaking country. Therefore, this factor was interpreted as indexing *L2 Nonacculturation*.

The eighth factor included learning L1 from radio and TV (positive loadings, Cronbach's α = .77) and was interpreted to reflect *Media-Based L1 Learning*.

Establishing predictive relationships using multiple regressions. Regression analyses are reported in Table 7, together with regression coefficients (marking the relative importance of each independent variable that entered the model) and statistics describing the fit of the model (B and β coefficients, VIF values, R and R^2 values, and F tests). VIFs associated with each independent variable ranged from 1.00 to 2.46, suggesting that no multicollinearity or singularity problems were present.

The language history measures that predicted self-reported proficiency in understanding, speaking, and reading are reported in the top panel of Table 7, self-reported proficiency measures that predicted behavioral performance are reported in the second panel of Table 7, language history measures that predicted behavioral proficiency are reported in the third panel of Table 7, and behavioral measures that predicted self-reported proficiency are reported in the bottom panel of Table 7.

Correlations between behavioral and self-reported measures. To establish criterion-based validity of self-reported proficiency measures, Pearson's R correlation analyses between self-reported and behavioral proficiency measures were conducted within each language and processing modality (see Table 8). The results yielded strong positive correlations between standardized behavioral measures (i.e., reading fluency, passage comprehension, productive vocabulary, oral comprehension, and grammaticality judgments) and self-reported measures of understanding, speaking, and reading L1 and L2. Although performance on standardized measures of sound awareness and receptive vocabulary did not relate to self-reported L1 proficiency, it was significantly related to self-reported L2 proficiency. The majority of standardized measures correlated more strongly with self-reported L2 proficiency than with self-reported L1 proficiency (with the exception of grammaticality judgment latencies, which correlated stronger with self-reported L1 proficiency). For L1, self-reported proficiency measures correlated most strongly with standardized behavioral measures of oral comprehension. For L2, the highest correlation values were obtained for passage comprehension and oral comprehension.

Discussion

The results of Study 2 confirmed questionnaire-based predictors of self-reported proficiency, identified questionnaire-based predictors of behavioral language performance, and revealed behavioral predictors of

Table 7. Multiple regression analyses for Study 2: Language history predictors of self-reported proficiency, self-reported proficiency predictors of behavioral performance, language history predictors of behavioral performance, and behavioral predictors of self-reported proficiency.

Predictee, *F* test	Predictor	Regression coefficients				Fit of model	
		B	*SE* of B	β	VIF	*R*	R^2
Language history predictors of self-reported proficiency							
Comprehending L1	Age began reading L1	−0.20	0.06	−.37	1.00	.37	.13
$F(1, 50) = 7.10, p < .05$							
Speaking L1	Age began reading L1	−0.28	0.06	−.54	1.05	.39	.39
$F(2, 50) = 24.58, p < .001$		0.13	0.04	.38	1.05	.72	.52
Reading L1	Exposure to L1 reading	0.18	0.05	.48	1.00	.48	.23
$F(1, 50) = 13.57, p < .001$							
Comprehending L2	Exposure to L2 friends	0.44	0.07	.68	1.01	.72	.52
$F(3, 50) = 21.6, p < .001$	Learning from L2 family	0.20	0.07	.24	1.10	.78	.61
	Age began acquiring L2	−0.10	0.05	−.22	1.11	.81	.52
Speaking L2	Exposure to L2 friends	0.39	0.06	.71	1.00	.72	.52
$F(2, 50) = 29.4, p < .001$	Learning from L2 family	0.17	0.06	.31	1.00	.79	.62
Reading L2	Exposure to L2 reading	0.28	0.09	.50	2.15	.74	.54
$F(2, 50) = 26.02, p < .001$	Exposure to L2 friends	0.16	0.08	.33	2.15	.77	.59
Self-reported proficiency predictors of behavioral performance							
L1 Reading Fluency (WJ/M)	Proficiency reading L1	9.24	2.91	.42	1.00	.42	.17
$F(1, 49) = 10.10, p < .01$							
L1 Passage Comprehension (WJ/M)	Proficiency reading L1	10.77	2.56	.52	1.00	.52	.27
$F(1, 49) = 17.78, p < .001$							
L1 Productive Vocabulary (WJ/M)	Proficiency reading L1	11.02	2.80	.50	1.00	.50	.25
$F(1, 49) = 15.55, p < .001$							
L1 Oral Comprehension (WJ/M)	Proficiency reading L1	11.01	1.80	.66	1.00	.66	.44
$F(1, 49) = 37.29, p < .001$							
L1 Grammaticality Accuracy	Proficiency reading L1	0.06	0.01	.52	1.00	.52	.27
$F(1, 49) = 17.17, p < .001$							
L1 Grammaticality Latency	Proficiency reading L1	452.5	98.87	−.56	1.00	.56	.31
$F(1, 49) = 20.94, p < .001$							
L1 Receptive Vocabulary (PPVT/TVIP)	Proficiency comprehending L1	5.42	2.43	.31	1.00	.31	.09
$F(1, 49) = 4.99, p < .05$							
L2 Reading Fluency (WJ/M)	Proficiency speaking L2	8.47	1.46	.64	1.00	.64	.41
$F(1, 49) = 33.62, p < .001$							
L2 Passage Comprehension (WJ/M)	Proficiency speaking L2	11.19	1.46	.74	1.00	.74	.55
$F(1, 49) = 58.40, p < .001$							
L2 Productive Vocabulary (WJ/M)	Proficiency speaking L2	7.14	1.24	.64	1.00	.64	.41
$F(1, 49) = 33.38, p < .001$							
L2 Oral Comprehension (WJ/M)	Proficiency speaking L2	10.72	1.44	.74	1.00	.74	.55
$F(1, 49) = 55.46, p < .001$							
L2 Grammaticality Accuracy	Proficiency speaking L2	0.05	0.01	.67	1.00	.67	.44
$F(1, 49) = 37.58, p < .001$							
L2 Grammaticality Latency							
$F(1, 49) = 7.79, p < .01$	Proficiency comprehending L2	−231.92	83.11	−.38	1.00	.38	.14
L2 Receptive Vocabulary	Proficiency speaking L2	8.32	1.44	.64	1.00	.64	.41
$F(1, 49) = 33.32, p < .001$							
L2 Sound Awareness (WJ/M)[a]	Proficiency reading L2	6.44	1.68	.53	1.00	.53	.28
$F(1, 49) = 14.74, p < .001$							
Language history predictors of behavioral proficiency							
L1 Reading Fluency (WJ/M)	Years in L1 country	1.26	0.34	.48	1.00	.48	.23
$F(1, 50) = 13.7, p < .001$							
L1 Passage Comprehension (WJ/M)	Years in L1 country	1.28	0.28	.53	1.07	.60	.36
$F(1, 50) = 17.3, p < .001$	Age attained L1 fluency	−2.6	1.06	−.28	1.07	.66	.44
L1 Productive Vocabulary (WJ/M)	General L1 exposure	0.29	0.09	.36	1.18	.53	.28

(Continued on the following page)

Table 7 *Continued.* Multiple regression analyses for Study 2: Language history predictors of self-reported proficiency, self-reported proficiency predictors of behavioral performance, language history predictors of behavioral performance, and behavioral predictors of self-reported proficiency.

Predictee, *F* test	Predictor	Regression coefficients				Fit of model	
		B	*SE* of B	β	VIF	*R*	R^2
$F(1, 50) = 10.7$, $p < .001$	Years in L1 schools	1.91	0.51	.58	2.07	.62	.38
	Years in L1 family	−1.26	0.52	−.36	1.94	.66	.43
	Learning L1 from tapes	−3.84	1.39	−.31	1.10	.71	.50
L1 Oral Comprehension (WJ/M)	Years in an L1 country	0.67	0.28	.34	1.20	.46	.21
$F(1, 50) = 8.9$, $p < .001$	Exposure to L1 reading	1.90	0.87	.30	1.20	.53	.28
L1 Grammaticality Accuracy	Exposure to L1 reading	0.02	0.01	.34	2.46	.75	.56
$F(1, 50) = 30.5$, $p < .001$	Learning L1 from reading	0.02	0.01	.28	1.42	.79	.63
	L1 general exposure	0.001	0.001	.35	2.23	.83	.68
L1 Grammaticality Latency	Exposure to L1 friends	−134.3	37.04	−.46	1.25	.60	.63
$F(1, 50) = 17.8$, $p < .001$	Learning L1 from reading	−120.4	46.29	−.33	1.25	.67	.45
L1 Receptive Vocabulary	Learning L1 from TV	−2.02	0.76	−.35	1.00	.36	.13
$F(1, 50) = 5.8$, $p < .01$	Learning L1 from tapes	−1.92	0.94	−.27	1.00	.45	.21
L1 Sound Awareness	Years in L1 schools	0.75	0.35	.33	1.00	.33	.11
$F(1, 50) = 4.6$, $p < .05$							
L2 Reading Fluency (WJ/M)	General L2 exposure	0.57	0.08	.76	1.04	.71	.51
$F(2, 50) = 23.7$, $p < .001$	L2 learning from radio	−2.06	0.90	−.26	1.04	.75	.57
L2 Passage Comprehension (WJ/M)	Exposure to L2 friends	4.75	1.04	.56	1.17	.68	.46
$F(1, 50) = 20.8$, $p < .001$	Years in an L2 family	0.83	0.34	.30	1.17	.73	.54
L2 Productive Vocabulary (WJ/M)	Exposure to L2 friends	3.39	0.70	.57	1.14	.68	.47
$F(2, 50) = 22.9$, $p < .001$	Age attained L2 fluency	−1.03	0.37	−.33	1.14	.75	.56
L2 Oral Comprehension (WJ/M)	General L2 exposure	0.38	0.12	.45	1.29	.63	.40
$F(1, 50) = 18.1$, $p < .001$	Years in an L2 family	0.10	0.35	.39	1.29	.72	.52
L2 Grammaticality Accuracy	Exposure to L2 TV	0.02	0.004	.51	1.77	.82	.67
	General L2 exposure	0.002	0.000	.37	1.77	.88	.77
$F(1, 50) = 50.4$, $p < .001$	Age attained L2 fluency	0.01	0.002	−.23	1.13	.90	.82
L2 Grammaticality Latency	Age attained L2 fluency	81.47	30.93	.38	1.23	.54	.29
$F(2, 50) = 11.49$, $p < .001$	L2 exposure (% time)	−15.23	6.17	−.36	1.23	.63	.40
L2 Receptive Vocabulary	Age started reading L2	−1.79	0.69	−.37	1.00	.39	.15
$F(2, 50) = 6.7$, $p < .01$	Learning L2 via friends	2.63	1.08	.35	1.00	.52	.27
L2 Sound Awareness	Years in an L2 country	1.08	0.22	.64	1.04	.58	.34
$F(3, 50) = 13.15$, $p < .001$	Learning L2 from radio	−3.31	0.89	−.47	1.03	.74	.54
	L2 self-instruction	−2.32	1.10	−.27	1.01	.78	.61
Behavioral predictors of self-reported proficiency							
Comprehending L1	Grammaticality latency	0.00	0.00	−.48	1.00	.48	.23
$F(1, 50) = 11.5$, $p < .01$							
Proficiency Speaking L1	Oral comprehension	0.04	0.01	.68	1.25	.54	.30
$F(2, 50) = 11.10$, $p < .001$	Sound awareness	−0.02	0.01	−.31	1.25	.61	0.37
Reading L1	Oral comprehension	0.03	0.01	.44	1.59	.64	.40
$F(2, 50) = 16.18$, $p < .001$	Grammaticality latency	0.00	0.00	−.32	1.59	.68	.47
Comprehending L2	Passage comprehension	0.04	0.01	.60	1.24	.72	.52
$F(2, 50) = 24.2$, $p < .001$	Grammaticality latency	0.00	0.00	−.28	1.24	.76	.58
Speaking L2	Passage comprehension	0.03	0.01	.42	2.58	.78	.61
$F(3, 50) = 26.22$, $p < .001$	Grammaticality accuracy	3.74	1.47	.31	1.66	.81	.66
	Receptive vocabulary	0.02	0.01	.26	1.78	.84	.70
Reading L2	Passage comprehension	0.03	0.01	.45	1.64	.67	.45
$F(1, 50) = 29.69$, $p < .001$	Grammaticality accuracy	3.73	1.57	.35	1.64	.73	.53

Note. WJ/M = Woodcock–Johnson Tests of Achievement/Woodcock–Muñoz Tests of Achievement.

[a]None of the self-reported proficiency measures predicted performance on the Sound Awareness task.

Table 8. Self-reported and behavioral proficiency correlations in Study 2.

Self-reported measures	Standardized behavioral measures in the same language							
	Reading fluency	Passage comprehension	Productive vocabulary	Oral comprehension	Sound awareness	Receptive vocabulary	Grammaticality judgments (Acc)	Grammaticality judgment (RT)
L1 Proficiency								
Speaking	.409**	.441**	.448**	.541**	0.008	0.236	.377**	−.434**
Comprehension	.346*	.428**	.413**	.481**	−0.025	.307*	.337*	−.432**
Reading	.417**	.520**	.495**	.661**	0.190	0.179	.517**	−.555**
L2 Proficiency								
Speaking	.642**	.741**	.640**	.739**	.499**	.640**	.667**	−.339*
Comprehension	.542**	.562**	.535**	.621**	.466**	.494**	.502**	−.377**
Reading	.532**	.634**	.586**	.575**	.529**	.574**	.603**	−.286*

Note. Acc = accuracy; RT = reaction time.

*$p < .05$. **$p < .01$.

self-reported language proficiency. Similar to participants in Study 1, participants in Study 2 reported high levels of proficiency and extensive immersion in both languages. Participants in Study 2 ranked exposure to an L1-speaking family as the most significant contributor to learning L1 and ranked reading in L2 as the most significant contributor to learning L2. Consistent with these rankings, participants reported that their highest L1 proficiency was in understanding, whereas their highest L2 proficiency was in reading. Participants' reports for L2 proficiency were lower than their reports for L1 proficiency, with the largest differences in understanding and the smallest differences in reading. Differences between self-reported proficiency levels in L1 and L2 were reflected in participants' performance on behavioral measures of linguistic ability, with significantly lower performance on standardized proficiency measures in L2 than in L1. For L2, participants reported high levels of reading proficiency while at the same time showing low performance on reading fluency and sound awareness tasks (sound awareness reflects phonological skills underlying reading). The discrepancy between self-reported and behavioral measures of reading suggests that bilinguals may overestimate their true L2 reading abilities. This tendency to overestimate L2 reading skill may be due in part to lack of direct feedback on performance during reading, which by its nature is less interactive than speaking or listening. The finding that bilinguals overestimate their skill in a weaker language is also consistent with learning research reporting that students overestimate their academic abilities when their performance is poor (Orsmond, Merry, & Reiling, 1997).

In Study 2, questions that were expected to measure the same underlying construct (i.e., language proficiency, etc.) clustered together in the factor analysis. Four factors (Relative L2–L1 Competence, L1 Learning, Late L2 Learning, and L1 Nondominant Status) accounted for more than half of the variance in the data. The remaining factors (L2 Immersion, L2 Nonacculturation, and Media-Based L1 Learning) accounted for less variance and may be characteristic of bilingual subgroups. Unlike L1 and L2 Competence in Study 1, L1 and L2 Competence in Study 2 did not cluster as separate factors. However, within the same factor, L1 and L2 Competence variables generated distinct patterns, with positive loadings for L2 variables and negative loadings for L1 variables. These findings replicate the central nature of L1 and L2 competencies in defining bilingual profiles and suggest that the two competencies may be interconnected depending on sample characteristics. Moreover, high Cronbach's alphas suggested that the majority of factors had high internal consistency. The four factors with highest Cronbach's alphas were Relative L2–L1 Competence, Late L2 Learning, L2 Immersion, and Media-Based L1 Learning. Cronbach's alpha values were somewhat lower for L1 Immersion and L1 Learning. Thus, variables describing Relative L2–L1 Competence and Late L2 Learning had high eigenvalues and high alpha scores, suggesting that they accounted for extensive variance and had high internal consistency. L2 Immersion and Media-Based L1 Learning had high internal consistency but accounted for relatively little variance in the data set.

Multiple regression analyses were used to generate predictive relationships between language history variables and self-reported proficiency for L1 and L2. L1 proficiency was predicted by age when participants began reading in L1 and by exposure to L1 reading. L2 proficiency was predicted by exposure to L2-speaking friends and learning L2 from family. The variables with the highest predictive values for L1 proficiency were the ones that participants reported as most relevant to learning L2, and the variables with highest predictive values for L2 proficiency were the ones that participants reported as most relevant to learning L1. In other words, although L2 reading played an important role in acquiring L2 and consistently emerged as a contributor to L2 learning, exposure to L2-speaking friends and family were the variables that were likely to vary the most in L2 and to therefore carry the most predictive value. The opposite pattern was true for L1, for which exposure to L1-speaking family and friends constituted a reliable factor in L1 learning and carried low predictive value for L1 proficiency, whereas reading-related variables were likely to vary the most in L1 and to therefore carry more predictive value.

Study 2 tested the questionnaire's criterion-based validity and established that self-reported measures of language status are indicative of bilinguals' linguistic performance. Results showed that proficiency in reading L1 was the best predictor of performance on standardized L1 tests and that proficiency in speaking L2 was the best predictor of performance on standardized L2 tests. This suggests that different self-reported proficiency measures were indicative of performance in the two languages, where the best predictors of actual performance were the dimensions on which bilinguals rated themselves the lowest. Of all behavioral measures administered in the two languages, only L1 sound awareness did not have a reliable predictor among self-reported proficiency measures. This may be because L1 sound awareness is a product of early language learning and could be more indicative of perceptual processing than of higher order language ability. Completing the sound awareness task in L1 is considerably easier (relative to L2), resulting in uniformly high scores and limited variability. Because data were analyzed in terms of L1 versus L2 (irrespective of whether L2 was Spanish or English), the finding that L2 (but not L1) sound awareness could not be predicted by self-reported proficiency measures suggests a difference between native versus second languages rather than a difference in specific language (Spanish or English) or in phonemic structures of the two languages. Study 2 also examined behavioral predictors of self-reported proficiency in L1 and L2 to identify the specific linguistic skills that bilinguals took into account when rating proficiency. The results suggest that measures of comprehension and judgments of grammaticality were the best predictors of self-reported proficiency in both languages. This finding suggests that reading comprehension and morphosyntactic abilities are core language skills that bilinguals take into consideration when judging language proficiency. It is possible that metacognitive awareness of these skills is especially high. Consistent with previous studies (e.g., Lemmon & Goggin, 1989; Ross, 1998), the observed correlations between self-reported and behavioral skills indicated that bilinguals' proficiency ratings corresponded to objective measures of language performance. Correlations between standardized and self-reported measures were stronger for L2 than for L1, likely because of a larger range of reported abilities in L2. A larger range of values yielded by L2 data resulted in greater variability to be accounted for by independent variables. As is inherent in statistical analyses, independent variables are better at capturing dependent variables with high variability, a pattern typically present in statistical comparisons of L2 variables relative to L1 variables (e.g., Flege et al., 2002).

Study 2 suggested that language history profiles underlying behavioral performance vary across tasks (e.g., receptive, productive, auditory, visual, single word, sentence level) and languages (first or second). This variability suggests that to be able to predict behavioral performance on a specific task in L1 or L2 solely on the basis of language history, one would have to take into account a different combination of language history variables, as identified in the *Results* section of each study. It appears, therefore, that self-reported proficiency may be a more efficient way of indexing overall language performance, with language history information augmenting predictions of performance on specific linguistic tasks.

General Discussion

The objective of this project was to develop a valid and reliable tool for assessing language proficiency: the LEAP-Q. In Study 1, the internal validity of the questionnaire was examined using factor analysis and multiple regression analyses. In Study 2, the internal validity of the questionnaire was replicated with a different sample of bilingual speakers, and criterion-based validity was examined using multiple regression and correlation analyses against a battery of standardized behavioral measures. Bilinguals' self-reported language proficiency was compared to objective measures of linguistic ability in both languages. The results of the two studies suggest that the LEAP-Q is an effective, efficient, valid, and reliable tool for assessing bilingual language status.

Factor analyses across studies yielded similar factorial structures, and most factors with high eigenvalues were consistent across the two studies. Specifically, L1 Competence and L2 Competence (separate factors in Study 1 but part of a combined Relative L2–L1 Competence factor in Study 2), as well as Late L2 Learning, accounted for extensive variance and had high internal consistency in both studies (as indicated by high eigenvalues and Cronbach's alphas). L2 Immersion accounted for a moderate amount of variance, whereas Media-Based Learning/Media-Based L1 Learning accounted for less variance but was internally consistent in both studies. Consistent clustering of questions across the two studies suggests that constructs revealed by factor analyses are reliable components of bilinguals' linguistic abilities. Although it is yet premature to suggest that questions within these factors might be combined into subscales for assessing L1 and L2 competence in diverse bilingual populations, the patterns are promising. For example, studies that target bilinguals' L1 proficiency may include questions that contributed to L1 status factors (e.g., L1 Maintenance and Balanced Immersion in Study 1, and Relative L1–L2 Competence and L1 Immersion in Study 2). These questions, which were also most predictive of behavioral L1 measures in regression analyses, address years in an L1 country, L1 school, and L1 family; general L1 exposure; and exposure to L1 reading. Similarly, studies that target L2 immersion may include questions that contributed to L2 immersion factors, such as Late L2 Immersion in Study 1 and L2 Immersion in Study 2. These questions address years in an L2 workplace, L2 school, and L2 family; exposure to an L2 family; and learning from an L2 family.

Although subscales might be a useful way to assess specific subsamples of bilinguals in the future, developing such subscales was not the objective of this study. Instead, the goal was to create a comprehensive questionnaire that is applicable across a diverse group of bilingual and multilingual populations. Subscale construction assumes constant relationships between questionnaire items; however, Harley et al. (1990) demonstrated that the factorial structure of L2 proficiency varied across bilingual groups with different language experiences. Therefore, data collected in this study illustrate meaningful relationships between questions (hence validating them) but did not provide evidence for specific subscales. Future studies may focus on specific subsamples of bilinguals and administer questions that clustered together in L1 and L2 Competence factors to examine their utility as scales. For instance, if variables that constituted L2 Competence factors in Studies 1 and 2 (e.g., length of exposure to L2, current exposure to L2 in different environments, proficiency speaking L2, etc.) formed a reliable subscale, then answers to these questions should reliably predict the level of L2 performance across a range of bilingual speakers.

A number of factors varied across studies; these included Non-Native Status, L1 Maintenance, and Balanced Immersion in Study 1 and L1 Nondominant Status, L1 Immersion, L1 Learning, and L2 Nonacculturation in Study 2. It appears that whereas L1 learning, immersion, and nondominant status clustered into three separate factors in Study 2, they were part of the combined L1 Maintenance factor in Study 1, which likely is due to differences in participants' characteristics. The finding that some factors were specific to either Study 1 or to Study 2 suggests that the LEAP-Q is versatile enough to reflect not only the general characteristics of bilingual experience (i.e., factors that converge across the two studies) but also population-specific characteristics of bilingual subsamples (i.e., factors that are idiosyncratic to each study).

In examining the external validity of the LEAP-Q, strong correlations between self-reported proficiency and standardized tests of linguistic performance were found, and preliminary predictive relationships between self-reported and standardized measures were established. The *fit* of the models, which refers to variance in the dependent variable accounted for by variability in the independent variable(s), averaged 62% for L2 and 50% for L1 and was comparable to values reported in previous research (e.g., MacIntyre et al., 1997). Self-reported proficiency in reading L1 was a reliable predictor of behavioral performance on standardized tests of language ability in L1. Self-reported proficiency in speaking L2 was a reliable predictor of behavioral performance on standardized tests of language ability in L2. We suggest that if questionnaire data are used to make general inferences about language function, then reading proficiency should be used to index L1, and speaking proficiency should be used to index L2. Because both simultaneous and sequential bilinguals were sampled in Study 2, it is likely that this observation will apply to various groups of bilingual individuals, with diverse acquisition histories. However, it is possible that this pattern of results, with speaking proficiency more indicative of L2 skills and with reading proficiency more indicative of L1 skills, is more applicable to sequential bilinguals (who acquired their first language early in life and their second language later in life). Future research may explicitly test the predictive power of self-assessed speaking and reading skills across various bilingual groups and directly compare sequential and simultaneous bilinguals.

Language history measures predicted self-reported and behavioral proficiency better for L2 than for L1. For instance, correlation values ranged from .5 to .68 for L1 and from .73 to .84 for L2 in analyses of self-reported proficiency, and from .45 to .83 for L1 and from .52 to .90 for L2 in analyses of behavioral performance. In Study 1, self-reported L1 proficiency was predicted by immersion in L1 environments, and self-reported L2 proficiency was predicted by L2 acquisition ages. In Study 2, self-reported L1 proficiency was predicted by age when participants began reading L1 and by exposure to L1 reading, whereas self-reported L2 proficiency was predicted by exposure to L2 friends and learning L2 from family. Despite the fact that the specific predictors of language proficiency varied across studies, a common pattern emerged; namely, in both studies, self-reported L1 proficiency was predicted by only a few language history attributes (e.g., age when participants began reading was a sole predictor for L1 proficiency understanding), whereas L2 proficiency levels were predicted by a more mixed set of predictors (e.g., exposure to L2-speaking friends, learning L2 from family, and age when participants began acquiring L2 were all predictors for L2 proficiency in reading). This finding may be related to the variability inherent in L2 learning compared with L1 learning.

In sum, similarities between factors in Studies 1 and 2 suggest that specific questions on the LEAP-Q cluster together in representing constructs of language experience and proficiency. Differences between factors and between predictive equations in Studies 1 and 2 suggest that the exact nature of the factors may be modulated by population characteristics. Strong correlations were found between self-reported and objective standardized proficiency measures, suggesting that self-reports are reliable indexes of language proficiency. Language history predictors of performance on standardized tasks provided a first step toward formulating specific hypotheses about the relationship between self-reported and objective language performance. In the future, regression coefficients using questionnaire-based information may be used for deriving estimates of behavioral performance. (Note that predictive equations for standardized ability in Study 2 were based on a relatively homogeneous Spanish–English/English–Spanish bilingual population. Therefore, these coefficients may be less useful for predicting language performance of individuals who come from a different population of bilinguals, and future work will need to examine predictive equations across bilingual populations.)

The results of Study 2 suggest that global measures of self-reported proficiency are good indicators of actual performance on specific measures of language ability; however, a precise estimate of performance on a specific task can be derived only using specific predictive equations that take into account multiple aspects of language history. This finding contributes to the debate about the utility of universal versus task-specific measures of language ability (e.g., Bahrick et al., 1994; Fishman & Cooper, 1969; Oyama, 1978; Spolsky et al., 1968) in that it suggests the value of both measures depending on the degree of precision desired. The LEAP-Q makes it possible to capture both global and detailed predictors of bilingual language performance. Thus, the LEAP-Q may enable use of language history information to predict actual language performance when direct measures of verbal behavior are not possible.

In the future, the LEAP-Q will be administered in experiments spanning learning in Spanish–English bilinguals, language and memory in Mandarin–English bilinguals, spoken language processing in German–English bilinguals, and written language processing in Russian–English bilinguals. In addition, the LEAP-Q is currently used during bilingual testing in several other laboratories. Thus, the next step will include cross-validation of the LEAP-Q by comparing self-reported language profiles to cognitive and behavioral measures across a wide range of experimental settings, including eye movement patterns, memory performance, and learning.

Use of the LEAP-Q

The LEAP-Q is intended for use with healthy adult bilinguals and multilinguals (both sequential and simultaneous) from diverse linguistic and cultural backgrounds who have attained high school levels of literacy. In its current form, the questionnaire cannot be used to index child language or speech-language disorders, and its extension to clinical populations has not been studied. At most, clinical applications of the questionnaire may include use with healthy bilingual/multilingual clients seeking accent reduction or to complement existing family reports. The LEAP-Q may also be used to obtain a more comprehensive account of clients' language histories for descriptive purposes, filling a need for gathering background information on bilingual/multilingual clients and providing a set of questions with proven relevance in nondisordered groups of similar language backgrounds.

The LEAP-Q is also available in an electronic format and can be obtained without charge by contacting Viorica Marian or by downloading it from the laboratory Web site (www.communication.northwestern.edu/csd/research/bilingualism). The questionnaire is currently available in English only (note that Delgado et al. [1999] showed that language of self-assessment does not influence bilinguals' proficiency ratings). The questionnaire takes approximately 15 min to complete for speakers of two languages (the time increases in 5-min increments for each additional language). It provides an extensive array of measures and can be completed by the bilingual independently (including before his or her arrival at the testing site), making it an effective addition to comprehensive bilingual assessment.

Acknowledgments

This work was supported by Grant 1R03HD046952-01A1 from the National Institute on Deafness and Other Communication Disorders and Grant BCS-0418495 from the National Science Foundation to Viorica Marian. We thank Dorothea Blumenfeld, Matthias Blumenfeld, Olga Boukrina, Joanna Bovee, Vridhi Chhabria, Nadia Cone, Caitlin Fausey, Dean Garstecki, Dong Lu, Naveen Malik, Avital Rabin, and Evar Strid for their contributions to this project.

References

Abdel-Khalek, A. M., Tomas-Sabado, J., & Gomez-Benito, J. (2004). Psychometric parameters of S-KUAS. *European Journal of Psychological Assessment, 20,* 349–357.

Bachman, L. F. (1990). Constructing measures and measuring constructs. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 26–49). Cambridge, England: Cambridge University Press.

Bachman, L., & Palmar, A. (1985). *Basic concerns in language test validation*. Reading, MA: Addison Wesley.

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General, 123,* 264–283.

Bedoya, E., Braun, G., Craig, L., Rodo, C., Senerth, D., & Hughes, E. R. (2005). *The best test preparation for the AP Spanish language exam* (5th ed.). Piscataway, NJ: Researchers & Education Association.

Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second language acquisition*. New York: HarperCollins.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161–181). Mahwah, NJ: Erlbaum.

Birdsong, D. (2005). Interpreting age effects in 2nd language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism* (pp. 109–127). New York: Oxford University Press.

Carroll, J. (1967). *The foreign language attainments of language majors in the senior year: A survey conducted in U.S. colleges and universities*. Cambridge, MA: Harvard University Press.

Chapnik-Smith, M. (1997). How do bilinguals access lexical information? In A. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 145–168). Mahwah, NJ: Erlbaum.

Chen, H. C. (1992). Lexical processing in bilingual or multilingual speakers. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 253–264). Amsterdam: Elsevier.

Chincotta, D., & Underwood, G. (1998). Non temporal determinants of bilingual memory capacity: The role of long-term representations and fluency. *Bilingualism, 1,* 117–130.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics, 20,* 1–20.

de Houwer, A. (1995). Bilingual language acquisition. In P. Fletcher & B. MacWinney (Eds.), *The handbook of child language* (pp. 219–250). London: Basil Blackwell.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22,* 499–533.

Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences, 21,* 31–46.

Demorest, M. E., & Erdman, S. A. (1987). Development of the Communication Profile for the Hearing Impaired. *Journal of Speech and Hearing Disorders, 52,* 129–143.

DeVellis, R. F. (1991). Scale development: Theory and applications. *Applied Social Research Methods Series, 26*. London: Sage.

Doctor, E., & Klein, D. (1992). Phonological processing in bilingual word recognition. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 237–252). Amsterdam: Elsevier.

Dörnyei, Z. (2003). *Questionnaires in second language research*. Mahwah, NJ: Erlbaum.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: AGS.

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody* (Adaptación Hispanoamericana) [Peabody Picture Vocabulary Test Hispanic American version]. Circle Pines, MN: AGS.

Favreau, M., & Segalowitz, N. (1982). Second language reading in fluent bilinguals. *Applied Psycholinguistics, 3,* 329–341.

Ferrer, M., Cordoba, J., Garin, O., Olive, G., Flavia, M., Vargas, V., et al. (2006). Validity of the Spanish version of the Chronic Liver Disease Questionnaire as standard outcome for quality of life assessment. *Liver Transplantation, 12,* 95–104.

Fishman, J., & Cooper, R. (1969). Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior, 8,* 276–282.

Flege, J., Frieda, A., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25,* 169–186.

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition, 20,* 155–187.

Flege, J. E., MacKay, I. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23,* 567–598.

Flege, J. E., Yeni-Komishian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41,* 78–104.

Genesee, F. (1985). Second language learning through immersion: A review of U.S. programs. *Review of Educational Research, 55,* 541–561.

Goggin, J. P., Estrada, P., & Villarreal, R. P. (1994). Picture-naming agreement in monolinguals and bilinguals. *Applied Psycholinguistics, 15,* 177–193.

Grainger, J. (1993). Visual word recognition in bilinguals. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 11–26). Amsterdam: John Benjamins.

Grosjean, F. (1982). *Life with two languages*. Cambridge, MA: Harvard University Press.

Grosjean, F. (2004). Studying bilinguals: Methodological and conceptual issues. In T. Bhatia & W. Ritchie (Eds.), *The handbook of bilingualism* (pp. 32–63). Malden, MA: Blackwell.

Harley, B., Cummins, J., Swain, M., & Allen, P. (1990). The nature of language proficiency. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 7–25). Cambridge, England: Cambridge University Press.

Harris, R. J., & McGhee-Nelson, E. M. (1992). Bilingualism: Not the exception any more. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 3–14). Amsterdam: Elsevier.

Hazan, V., & Boulakia, G. (1993). Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech, 36,* 7–38.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in second language acquisition. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics, 23,* 599–621.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kim, K. H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997, July 10). Distinct cortical areas associated with native and second languages. *Nature, 388,* 171–174.

Kohnert, K., Bates, E., & Hernandez, A. (1999). Balancing bilinguals: Lexical–semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research, 42,* 1400–1413.

Kohnert, K., Kennedy, M. R. T., Glaze, L., Kan, P. F., & Carney, E. (2003). Breadth and depth of diversity in Minnesota: Challenges in clinical service competency. *American Journal of Speech-Language Pathology, 12,* 259–272.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. de Groot & J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Mahwah, NJ: Erlbaum.

Kuk, F. K., Tyler, R. S., Russel, D., & Jordan, H. (1990). The psychometric properties of a tinnitus handicap questionnaire. *Ear and Hearing, 11,* 434–445.

Lambert, W. (1955). Measurement of the linguistic dominance of bilinguals. *Journals of Abnormal Social Psychology, 50,* 197–200.

Lemmon, C. R., & Goggin, J. P. (1989). The measurement of bilingualism and its relationship to cognitive ability. *Applied Psycholinguistics, 10,* 133–155.

MacIntyre, P. D., Noels, K. A., & Clement, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47,* 265–287.

MacKay, I., & Flege, J. (2004). Effects of the age of second language learning on the duration of first and second language sentences. *Applied Psycholinguistics, 25,* 373–396.

Macnamara, J. (1969). How can one measure the extent of a person's bilingual proficiency? In L. Kelly (Ed.), *Description and measurement of bilingualism* (pp. 80–119). Toronto, Ontario, Canada: University of Toronto Press.

Macnamara, J., & Kushnir, S. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior, 10,* 480–487.

Magiste, E. (1979). The competing language systems of the multilingual. *Journal of Verbal Learning and Verbal Behavior, 18,* 79–89.

Magiste, E. (1992). Second language learning in elementary and high school students. *European Journal of Cognitive Psychology, 4,* 355–365.

Marian, V. (in press). Bilingual research methods. In J. Altarriba & R. R. Heredia (Eds.), *An introduction to bilingualism: Principles and processes*. Mahwah, NJ: Erlbaum.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2003, May). *Developing a language experience and bilingual status questionnaire*. Poster presented at the annual meeting of the Midwestern Psychological Association, Chicago.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129,* 361–368.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing. *Bilingualism: Language and Cognition, 6,* 97–115.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86,* 70–82.

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics, 21,* 395–423.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Mendenhall, W., & Sincich, T. (1996). *A second course in statistics: Regression analysis*. Englewood Cliffs, NJ: Prentice Hall.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling, 2,* 255–273.

Muñoz-Sandoval, A. F., Woodcock, R. F., McGrew, K. S., & Mather, N. (2005). *Pruebas de aprovechamiento* [Tests of achievement] (Batería III, Woodcock–Muñoz). Itasca, IL: Riverside.

Orsmond, P., Merry, S., & Reiling, K. (1997). Tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education, 22,* 357–369.

Ortiz, A., & Garcia, S. (1990). Using language assessment data for language and instructional planning for exceptional bilingual students. In A. Carrasquillo & R. Baecher (Eds.), *Teaching the bilingual special education student* (pp. 25–47). Norwood, NJ: Ablex.

Oyama, S. (1978). The sensitive period and comprehension of speech. *Working Papers on Bilingualism, 16,* 1–17.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121,* 1841–1852.

Romaine, S. (1995). *Bilingualism*. Oxford, England: Blackwell.

Roseberry-McKibbin, C., Brice, A., & O'Hanlon, L. (2005). Serving English language learners in school settings: A national survey. *Language, Speech, and Hearing Services in Schools, 36,* 48–61.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 5,* 1–20.

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing, 15,* 86–108.

Snow, C. E. (1983). Age differences in second language acquisition: Research findings and folk psychology. In K. M. Bailey, M. Long, & S. Peck (Eds.), *Second language acquisition studies* (pp. 141–150). Rowley, MA: Newbury House.

Snow, S., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49,* 1114–1128.

Snow, C. E., & Hoefnagel-Höhle, M. (1979). Individual differences in second language learning ability: A factor analytic study. *Language and Speech, 22,* 1515–1562.

Spolsky, B., Sigurd, B., Sako, M., Walker, E., & Arterburn, C. (1968). Preliminary studies in the development of techniques for testing overall second language proficiency. *Language Learning, 18,* 79–101.

Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19,* 69–75.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Talamas, A., Kroll, J., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2,* 45–58.

Tapia, M. (2001). Measuring emotional intelligence. *Psychological Reports, 88,* 353–364.

Tran, T. V. (1994). Bilingualism and subjective well-being in a sample of elderly Hispanics. *Journal of Social Service Research, 20*(1–2), 1–19.

U.S. Bureau of the Census. (2003). *Language use and English-speaking ability* (Census 2000 Brief C2KBR-29]. Retrieved November 24, 2004, from www.census.gov/population/www/cen2000/briefs.html.

Vaid, J., & Hull, R. (2002). Re-envisioning the bilingual brain using fMRI. In F. Fabbro (Ed.), *Advances in the neurolinguistics of bilingualism* (pp. 315–355). Udine, Italy: Forum.

Vaid, J., & Menon, R. (2000). Correlates of bilinguals' preferred language for mental computations. *Spanish Applied Linguistics, 4,* 325–342.

Weber-Fox, C., & Neville, H. J. (1999). Functional neural subsystems are differentially affected by delays in 2nd language immersion. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 23–38). Mahwah, NJ: Erlbaum.

Wiebe, J. S., & Penley, J. A. (2005). A psychometric comparison of the Beck Depression Inventory—II in English and Spanish. *Psychological Assessment, 17,* 481–485.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson Tests of Cognitive Abilities* (3rd ed.). Itasca, IL: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, F. (1993). *Woodcock–Johnson Language Survey.* Itasca, IL: Riverside.

Wu, Y., Larrabee, J. H., & Putman, H. P. (2006). Caring behaviors inventory—A reduction of the 42-item instrument. *Nursing Research, 55,* 18–25.

Zea, M., Asner-Self, K., Birman, D., & Buki, L. (2003). The Abbreviated Multidimensional Acculturation Scale. *Cultural Diversity and Ethnic Minority Psychology, 9,* 107–126.

Received March 13, 2006

Revision received July 25, 2006

Accepted October 2, 2006

DOI: 10.1044/1092-4388(2007/067)

Contact author: Viorica Marian, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208.
E-mail: v-marian@northwestern.edu.

Appendix (p. 1 of 2). Language Experience and Proficiency Questionnaire.

Last Name		First Name		Today's Date	
Age		Date of Birth		Male ☐	Female ☐

(1) Please list all the languages you know in order of dominance:

1 Language A	2 Language B	3 Language C	4 Language D	5 Language E

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1 Language A	2 Language B	3 Language C	4 Language D	5 Language E

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language.
(*Your percentages should add up to 100%*):

List language here:	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.
(*Your percentages should add up to 100%*):

List language here	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.
(*Your percentages should add up to 100%*):

List language here	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(6) Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc.):

List cultures here	Culture A (click here for scale)	Culture B (click here for scale)	Culture C (click here for scale)	Culture D (click here for scale)	Culture E (click here for scale)

(7) How many years of formal education do you have?________________________
Please check your highest education level (or the approximate U.S. equivalent to a degree obtained in another country):

☐ Less than High School
☐ High School
☐ Professional Training
☐ Some College
☐ College
☐ Some Graduate
☐ Masters
☐ PhD/MD/JD
☐ Other:

(8) Date of immigration to the United States, if applicable: ________________________
If you have ever lived in another country, please provide name of country and dates of residence:
____

(9) Have you ever had a vision problem ☐, hearing impairment ☐, language disability ☐, or learning disability ☐? (Check all applicable).
If yes, please explain (including any corrections): ________________________

Appendix (p. 2 of 2). Language Experience and Proficiency Questionnaire.

Language: Language X

This is my (please select from scroll-down menu: First, Second, Third, etc.) language.
All questions below refer to your knowledge of Language X.

(1) Age when you…:

began acquiring Language X:	*became fluent* in Language X:	*began reading* in Language X:	*became fluent reading* in Language X:

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where Language X is spoken		
A family where Language X is spoken		
A school and/or working environment where Language X is spoken		

(3) On a scale from zero to ten, please select your *level of proficiency* in speaking, understanding, and reading Language X from the scroll-down menus:

Speaking	(click here for scale)	Understand spoken language	(click here for scale)	Reading	(click here for scale)

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning Language X:

Interacting with friends	(click here for scale)	Language tapes/self instruction	(click here for scale)
Interacting with family	(click here for scale)	Watching TV	(click here for scale)
Reading	(click here for scale)	Listening to the radio	(click here for scale)

(5) Please rate to what extent you are currently exposed to Language X in the following contexts:

Interacting with friends	(click here for scale)	Listening to radio/music	(click here for scale)
Interacting with family	(click here for scale)	Reading	(click here for scale)
Watching TV	(click here for scale)	Language-lab/self-instruction	(click here for scale)

(6) In your perception, how much of a foreign accent do you have in Language X?
(click here for scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in Language X:
(click here for scale)