PROOFREADING INSTRUCTIONS

Dear Author,

Here is a proof of your article for publication in Applied Psycholinguistics. Please print out the file and check the proofs carefully, make any corrections necessary on a hard copy, and answer queries on the e proofs.

You may choose one of the following options for returning your proofs.

If the corrections can be explained clearly in a text message, please list the corrections in an email, citing page number, paragraph number, and line number. You may also list the corrections in a Doc file. Send the corrections to the Senior Project Manager, Nancy BriggsShearer, at this email address: <nbs@bsons.us>

If the corrections are extensive, please send the hard copy corrections no more than 48 hours after receipt to: Senior Project Managing Editor, Nancy L. BriggsShearer, 787 Moran Road, Oxford, New York 13830-3334, USA.

You are responsible for correcting your proofs. Errors not found may appear in the published journal.

The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted, unless discussed with the editor of the journal.

Please answer carefully any queries raised from the typesetter.

A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the typesetter is required—please provide this in electronic file form.

Thank you in advance.
Applied Psycholinguistics (APS)

Billing Address
Name: 
Company: 
Address: 
City: _________ State: _______ Zip: _______
Country: 
Phone: _________ Fax: _______

Copies

<table>
<thead>
<tr>
<th>No. of Pages</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>100</th>
<th>Add'l 50's</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>70</td>
<td>90</td>
<td>115</td>
<td>140</td>
<td>28</td>
</tr>
<tr>
<td>9-16</td>
<td>90</td>
<td>115</td>
<td>145</td>
<td>170</td>
<td>33</td>
</tr>
<tr>
<td>17-24</td>
<td>110</td>
<td>135</td>
<td>165</td>
<td>200</td>
<td>46</td>
</tr>
<tr>
<td>25-32</td>
<td>125</td>
<td>160</td>
<td>200</td>
<td>240</td>
<td>59</td>
</tr>
<tr>
<td>Add'l 18 pp</td>
<td>15</td>
<td>25</td>
<td>35</td>
<td>45</td>
<td>12</td>
</tr>
<tr>
<td>Covers Add'l</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>6</td>
</tr>
</tbody>
</table>

Article Reprint Being Ordered
Author(s): 
Article Title: 
Log No.: _________ Volume: _______ Issue: _______
Page Begin: _______ Page End: _______

*All Reprints are 8 1/2 x 11*

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints</td>
<td>$______</td>
</tr>
<tr>
<td>Covers</td>
<td>$______</td>
</tr>
<tr>
<td>Shipping Cost (see below)</td>
<td>$______</td>
</tr>
<tr>
<td>Total Amount Due</td>
<td>$______</td>
</tr>
</tbody>
</table>

Shipping Address
Name: 
Company: 
Address: 
City: _________ State: _______ Zip: _______
Country: 
Phone: 
Email: 

Method of Payment
*If using a credit card, please fill in the credit card billing address above*
- [ ] P.O. Number, (Attach a copy to the order form)
- [ ] Check or Money Order
  (U.S. Funds payable to Odyssey Press Inc.)
- [ ] AMEX  [ ] MasterCard  [ ] Visa
Card #: 
Expiration Date: 
Signature: 

Payment Terms
- Domestic - orders must be paid in advance
- United States Institution or Company - orders must be paid in advance or by purchase order
- International - orders must be paid in advance and the funds must be drawn on a U.S. Bank and payable in U.S. dollars to:
  Odyssey Press Inc.
  P.O. Box 7307
  Gonic, NH 03839-7307
  Fax: 1-603-822-2851

Shipping Information
- U.S. Domestic Orders - UPS Ground included
- U.S. Alaska & Hawaii - UPS Second Day Air
$40.00 per 100 reprints ($40.00 minimum)
- International Orders - Printed matter book
$40.00 per 100 reprints (3-6 month delivery)
- International Orders - Air printed matter
$105.00 per 100 reprints (2-4 week delivery)
EZ Reprint Ordering Instructions
*Please Read Carefully*

Please note that this form is being sent only to the author receiving page proofs. Accordingly, please be sure that all co-authors’ reprint orders are included on the attached form. When returning corrected page proofs to the publisher, please return reprint orders separately to the address below.

Odyssey Press Inc. will only process your reprint order if it is accompanied by the following:

- the completed EZ REPRINT ORDER FORM and...

  - For individual author and ALL International orders...
    - a valid Visa, MasterCard, American Express, check, or money order for the correct amount. The funds must be drawn on a U.S. bank and payable in U.S. dollars

  - For U.S. institutional orders...
    - an original purchase order, or a copy of the original purchase order

U.S. institutional orders not accompanied by a purchase order, will not be processed. If your institution does not issue purchase orders, than you must prepay by check or credit card.

Be sure that you complete the entire EZ REPRINT ORDER FORM legibly. Incomplete or illegible order forms will cause a delay in shipping your order. If you have received this as a pdf. file, please fill it out before printing.

Returned checks will be subject to a $25.00 reprocessing charge.

If your order is being shipped U.S. Alaska & Hawaii or outside the United States, you must complete the shipping instructions. The shipping charge must be included in the check/money order or credit card amount.

Reprints are shipped 30 days from publication date. Please do not call or write with inquiries until 30 days after publication.

For questions please E-mail: reprints@odysseypress.com

**MAIL OR FAX ORDER FORM TO:**
Odyssey Press Inc.
P.O. Box 7307
Gonic, NH 03839-7307

Fax: (603) 822-2851

*Fill out order form completely*
Author Queries Article APS 1000005

AQ1: Please confirm Bialystok 2002 reference.
AQ2: Please provide Lenneberg 1967 reference.
Cross-linguistic evidence for the nature of age effects in second language acquisition

ROBERT DEKEYSER
University of Maryland at College Park

IRIS ALFI-SHABTAY and DORIT RAVID
Tel-Aviv University

Received: February 18, 2008 Accepted for publication: May 18, 2009

ADDRESS FOR CORRESPONDENCE
Robert DeKeyser, Department of Linguistics, University of Maryland, 3215 Jimenez Hall, College Park, MD 20742. E-mail: rdk@umd.edu

ABSTRACT
Few researchers would doubt that the ultimate attainment in second language grammar is negatively correlated with age of acquisition, but considerable controversy remains about the nature of this relationship: the exact shape of the age-attainment function and its interpretation. This article presents two parallel studies with native speakers of Russian: one on the acquisition of English as a second language in North America (n = 76), and one on the acquisition of Hebrew as a second language in Israel (n = 64). Despite the very different nature of the languages being learned, the two studies show very similar results. When age at testing is partialled out, the data reveal a steep decline in the learning of grammar before age 18 in both groups, followed by an essentially horizontal slope until age 40. This is interpreted as evidence in favor of the critical period. Both groups show a significant correlation between ultimate attainment and verbal aptitude for the adult learners, but not for the early learners. This is interpreted as further evidence that the learning processes in childhood and adulthood not only yield different levels of proficiency but are also different in nature.

Age effects in (second) language learning are widely acknowledged, but their exact nature remains controversial, in particular, the concept of a critical period for second language acquisition (SLA). In the about the last 15 years, numerous arguments against the critical period hypothesis (CPH) have been formulated: a few studies have failed to find a clear correlation between age of acquisition and ultimate attainment; many more researchers accept the negative correlation as a fact, but they argue that it is attributable to a confound between age of acquisition and one or more other variables, such as length of residence, age at testing, the nature of the input received as a function of age, the extent to which education was provided in the second language (L2) or the first language (L1), the (lack of) motivation to integrate fully with the L2 society, or simply the amount of practice.
in the L2 as opposed to the L1. Some have also taken the rather different results found for learners of different L1s learning the same L2 as evidence that age of acquisition is not an important predictor by itself. (For recent critical overviews of the literature for and against the CPH, see Birdsong, 2005; DeKeyser & Larson-Hall, 2005; Herschensohn, 2007; Hyltenstam & Abrahamsson, 2003; Ioup, 2005; Long 2007, chap. 3.) The argumentation for and against the CPH that has received the most attention in the literature of the last 5 to 10 years, however, has been about the very nature of the age of acquisition–ultimate attainment function, which is centered on the question of whether the discontinuity in development implied by the CPH is found in the various data sets that were analyzed.

THE SHAPE OF THE AGE OF ACQUISITION–ULTIMATE ATTAINMENT FUNCTION

For about the last 10 years a number of researchers have analyzed data sets that appear to show a negative correlation between age of acquisition and ultimate attainment throughout the life span (Bialystok & Hakuta, 1999; Bialystok & Miller, 1999; Hakuta, Bialystok, & Wiley, 2003) or even a stronger negative correlation in adulthood than in childhood or adolescence (Birdsong & Molis, 2001). Therefore, these researchers argue, the decline as a function of age, which has been documented in dozens of studies, should not be interpreted as a critical period effect (a sharp decline within a relatively short period of time, attributable to inevitable biological and psychological changes, leveling off quickly once that period of time is over, and affecting only very particular kinds of learning, in this case the learning of aspects of a second language). It instead reflects a broader phenomenon of maturationally determined cognitive decline that is largely the same throughout the life span, but perhaps accentuated at certain stages of life by changes in people’s socialization patterns (see also Birdsong, 2004, 2005, 2006). For various methodological criticisms of these studies purporting to show a decline throughout adulthood, see DeKeyser (2006), Long (2005, 2007, chap. 3), and Stevens (2004).

Other researchers have countered that the well-documented decline as a function of age is a maturational phenomenon affecting (second) language learning more than other cognitive functions, and takes places in a period of roughly 10 to 15 years, starting possibly at birth, becoming clearly visible from around age 6 for certain aspects of language and with certain test formats, and leveling off in late adolescence. As evidence for this viewpoint, they point to qualitative differences in learning processes before and after this critical period, such as a differential role of aptitude (DeKeyser, 2000; Harley & Hart, 1997, 2002), or simply the shape of the age of acquisition–ultimate attainment function, which, at least in a number of studies, appears to show a steep decline during a limited number of years, but not thereafter (e.g., Flege, Yeni-Komshian, & Liu, 1999; Johnson, 1992; Johnson & Newport, 1989, 1991; Lee & Schachter, 1997; cf. DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003).

The issue remains far from resolved for a number of methodological reasons. The quantitative evidence (i.e., about how much of the L2 is learned as a function of age of acquisition) used on either side of the debate is often less than ideal,
because of very narrow operationalizations of ultimate attainment (most often grammaticality judgments) and insufficient documentation or analysis of potential confounding variables (especially length of residence, age at testing, the nature of the immigrants’ social networks in L1 and L2, the amount of use of L1 and L2 at various stages of development, and various affective and social–psychological variables).

The qualitative evidence (about how L2 learning takes place at different ages), in contrast, is rather limited so far (few researchers have focused on qualitative distinctions) and has been subject to a number of criticisms as well. DeKeyser and Larson-Hall (2005) mention differential reliance on aptitude at different ages and differential age effects on the learning of structures characterized by different levels of salience as potential examples of (somewhat indirect) evidence for qualitative age differences, that is, differences not just in learning outcomes, but also in learning mechanisms involved at different ages. Systematic research on the relationship between age and salience has not been carried out so far, however, and the finding of different predictive validity of aptitude at different ages (e.g., DeKeyser, 2000; Harley & Hart, 1997, 2002) has been questioned because of the instruments used. Bialystok, for instance, argued that the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959; used by DeKeyser, 2000) “aside from being almost 50 years old . . . investigates a narrow and almost parochial version of language aptitude” (2002, p. 484); presumably her criticisms would apply to the Pimsleur Language Aptitude Battery (PLAB; Pimsleur, 1966; used by Harley & Hart, 1997) as well, as this test was published only a few years after the MLAT and has largely the same ingredients and the same predictive validity (see, e.g., Carroll, 1981). Bialystok does not suggest any alternatives, however, nor is any test available at this point that is generally agreed to be a more valid measure of language learning aptitude than the MLAT or PLAB.1

In addition to issues of sampling and instrumentation, several further methodological problems cannot be ignored. One is that different aspects of language (e.g., phonology vs. morphosyntax, or even at a much more fine-grained level, regular inflection vs. irregular inflection) may show different age of acquisition–ultimate attainment functions. There may be “multiple windows” (multiple critical periods) for different aspects of language, some closing before others or showing a steeper decline than others (for early mentions of this idea, see, e.g., Schachter, 1996; Seliger, 1978); some aspects of language may also be more sensitive to variables such as length of residence or level of education than to age of acquisition (see especially Flege et al., 1999). Simply generalizing to all of “language” would be unfruitful (Eubank & Gregg, 1999, p. 66). In contrast, “it would be premature to exclude factors other than ‘linguistic competence’ from the agenda of maturational constraints” (Hyltenstam & Abrahamsson, 2003, p. 559); one may have to distinguish aspects such as the capacity “to construct narratives, to produce and understand metaphor, to accommodate to another’s speech, to persuade . . . ” (Schumann, 1995, p. 60). All of this also implies the necessity of cross-linguistic research, given that different native languages and target languages differ considerably in the frequency (or absence) of a wide variety of phenomena, such as phonemic use of tone, irregular inflectional morphology, agreement patterns within the noun phrase or between the noun phrase and the verb phrase, morphological...
marking of marked semantic distinctions, and wh-movement, to name just a few. Therefore, it is important to look at both the acquisition of the same L2 by speakers of very different L1s and the acquisition of very different L2s by speakers of the same L1. Thus far, only the former comparison has been made in the CPH literature on the acquisition of grammar (Bialystok & Miller, 1999; Hyltenstam, 1992; McDonald, 2000; Sorace, 1993).

Another problem is the inherent relationship between three important predictors of ultimate attainment—age of acquisition, length of residence, and age at testing. In most studies, there is a moderate to strong correlation between these three variables (age at testing = age of acquisition + length of residence, so if length or residence varies little, age of acquisition and age at testing will be strongly correlated; if age at testing varies little, age of acquisition and length or residence will be strongly correlated). Stevens (2006) argues that the linear dependence between age of acquisition, length of residence, and age at testing is very hard to disentangle, and can only be resolved through longitudinal data, or by measures of quantity of exposure not expressed in units of time, or by positing nonlinear relationships. Simply ignoring one of the three variables does not work, unless “two of the three variables can be regarded as indexing the same causal phenomena or if one of the variables is unrelated to the dependent variable” (p. 680). Length or residence is taken into account in most studies, and turns out to be unrelated to most dependent measures, provided that length of residence is more than 5 years, and that the dependent measures index basic grammatical proficiency (not purisms, collocations, etc.); it is therefore not much of a problem in most studies. The most problematic variable is age at testing, which is often not taken into account despite its sometimes high correlation with age of acquisition.

Finally, the evidence is only as good as the sample. Sample sizes in CPH studies, at least the ones focusing on the acquisition of grammar, have typically varied around 50, which is very small if the sample needs to be divided up into different age of acquisition ranges, and if a correlation, and especially a partial correlation or regression equation, needs to be computed for each subsample, as is typically the case. Of more importance, the qualitative nature of the sample, especially with respect to socioeconomic and educational diversity, monolingual or bi-/multilingual background, and relative size of different age of acquisition groups in the sample, often leaves much to be desired in terms of representativeness. In this area of research, almost every sample has been one of convenience, which typically means a much higher percentage of highly educated participants than in the population at large, and sometimes knowledge of one or more L2s before immigration. Moreover, a number of studies with an otherwise respectable number of participants have had few in the critical age range of 12–18; a few teenage participants more or less within the subsample of “early acquirers” (often defined as those who immigrated before age 16 or 18) may lead to large differences in the age of acquisition–ultimate attainment correlation for that sample, as the decline as a function of age is expected to be most noticeable between the ages of about 12 and 18, depending on the nature of the outcome variable and the L1–L2 difference. This alone may explain the rather large difference in the value for this correlation in studies with otherwise very similar results, such as Johnson and Newport (1989) and DeKeyser (2000).
From this short literature review, it is clear that, in order to make solid progress in investigating the CPH, we need to

1. distinguish morphosyntax from phonology or the lexicon, perhaps even different elements of morphosyntax;
2. conduct separate analyses (whether they be correlations, partial correlations, or regression analyses) per age group;
3. carry out studies with more subjects than has usually been the case, to ensure enough statistical power for these separate analyses;
4. conduct more research on qualitative age differences, for example, on whether aptitude plays a different role at different ages; and
5. design more cross-linguistic research for purposes of generalization. Only when learners with different L1 backgrounds learn the same L2, or when learners of the same L1 background learn different L2s and their data are collected and analyzed in the same way, preferably in the same study, is it possible to assess to what extent the nature of target language structures or the nature of L1–L2 differences interacts with age effects.

In this article, we report on a research project on the acquisition of L2 grammar that was designed to meet most of these goals. Data were collected from native speakers of Russian who acquired either English as an L2 in North America or Hebrew as an L2 in Israel (~150 participants). Results were analyzed separately for different ranges of age of acquisition, and the role of aptitude was investigated in each group to test for qualitative differences. As stated above, different aspects of language should all be investigated; this does not necessarily have to happen in one and the same study. It does seem important, however, to have a good sampling of one area, in this case morphosyntax, so that some generalization to at least that aspect of language is possible, but not to the lexicon, pragmatics, or pronunciation, of course.

Another limitation of this study is that no detailed information was collected about participants’ use of L1 and L2 from immigration to the time of testing. Although L1 and L2 are obviously in complementary distribution, their relative frequency tends to correlate with age of acquisition (see especially Bylund, 2008, in press; Jia & Aaronson, 2003; Montrul, 2008), and the degree of L2 acquisition tends to correlate with the degree of L1 attrition (see especially Köpke & Schmid, 2004; Schmid, 2006), it is virtually impossible to get good measures of quantity and quality of input from immigration to the time of testing without a longitudinal study (and a longitudinal study, in turn, is almost impossible to carry out with the number of subjects required for statistical reasons).

The following hypotheses were tested:

Hypothesis 1: For both the L2 English and the L2 Hebrew group, the slope of the age of arrival–ultimate attainment function will not be linear throughout the life span, but will instead show a marked flattening between adolescence and adulthood.
Hypothesis 2: The relationship between aptitude and ultimate attainment will differ markedly for the younger and older arrivals, with significance for the latter only.
These hypotheses require that a cutoff point be established between early and late acquirers. It also seems prudent to make a further distinction between relatively young adults and middle-aged acquirers (who, of course, by the time of testing, may already be senior citizens). Ideally, with an extremely large number of subjects, one could let any observed discontinuities in the age of acquisition–ultimate attainment function serve as cutoff points. In practice, however, the only alternative (cf. DeKeyser & Larson-Hall, 2005) is to choose empirically motivated cutoff points, even if these remain somewhat arbitrary. Although age 12 was often mentioned as a turning point in early literature (e.g., Lenneberg, 1967) and it has been used in some recent studies (e.g., Abrahamsson & Hyltenstam, 2008; McDonald, 2006), a number of studies by researchers with otherwise very different views on age issues, for example, Bialystok and Hakuta (1994), Johnson and Newport (1989, 1991), and DeKeyser (2000), show that the steep decline in ultimate attainment continues through adolescence. An arbitrary cutoff at age 12 would therefore seriously underestimate the age of arrival–ultimate attainment correlation among early learners and overestimate it among later learners. Furthermore, as middle-aged and senior citizens are likely to perform less well on a variety of tests for independent reasons, it is important that the results for participants over 50 years old at the time of data collection, that is, roughly over 40 years old at the time of immigration, are analyzed separately. In this study, we decided to analyze three separate data slices: <18, 18–40, and >40.

RATIONALE FOR THE CROSS-LINGUISTIC RESEARCH PROJECT

English and Hebrew are typologically very different languages, particularly in the area of morphology, and therefore ideally suited for this type of cross-linguistic research.

Hebrew is a Semitic language that is considered to be morphologically rich because it expresses many notions morphologically, and it offers a wide array of structural options to express these notions. Nouns and adjectives are obligatorily inflected for gender and number, for example,

ha-maxbar-ot the-notebook-s
ha-adum-ot the-red, Fm Pl
"the red notebooks"

Verbs are obligatorily inflected for gender, number, person, and tense in past and future tenses, for example,

ha-maxbér-et ne’elm-a
"the-notebook—Fm, 3rdSg. disappear-ed, Fm, 3rdSg."
"the notebook disappeared"

In present tense, verbs are inflected like adjectives and nouns. Prepositions, a closed-class category, incorporate pronominal information in their obligatory inflection for gender, number, and person, for example, lax “to-you, Fm.” In addition, numerals agree with nouns in gender, although the agreement system is
opaque and hard to learn (Ravid, 1995b). Finally, Hebrew marks gender, number, and person optionally on genitive nouns and accusative verbs side by side with syntactic constructions expressing the same notions.

Hebrew has two basic word orders: subject–verb–object (SVO), with either a lexical or pronominal subject and a lexical verb (or a copula), for example,

\[ \text{dan hevin et ha-inyan} \]

“Dan understood Acc the-matter”; and a predicate-first word order, expressing existence, possession, and modal meanings, typically containing a less “verbal” predicate and often subjectless, for example,

\[ \text{kday le-xa la-vo} \]

better to-you to-come

“you’d better come over” (Berman, 1980; Ravid, 1995a).

Word order is not rigid, given the rich agreement systems in Hebrew, which marks thematic and syntactic roles clearly and transparently.

In contrast, English is an SVO language with strongly grammaticized rather than pragmatically determined constraints on word order (Thompson, 1978). This is partially attributable to its impoverished system of grammatical inflection. Case is morphologically distinguished only in pronouns and in genitive phrases; subjects and direct objects occur as bare noun phrases with no overt case marking, whereas datives and oblique objects and adjuncts are marked by prepositions rather than by inflections.

Another facet of its lack of grammatical inflection is that English has almost no marking of agreement for gender, number, or person; the only exceptions are subject–verb concord with the verb be and third person present-tense marking by final -s. As a result, ordering of constituents is the major indicator of grammatical relations. Even postverbal elements are fairly strictly ordered, because nothing can be interspersed between the verb and its direct object, and locatives typically precede temporal adverbials (Berman, 1999). One clear exception to this morphological sparseness is that English marks comparative and superlative values on adjectives through morphology (e.g., slower, biggest), although this inflection is restricted to short adjectives, usually of Germanic origin.

Russian, the L1 in both studies, is a Slavic language with very rich inflectional morphology, but does not use articles. Nouns, adjectives, and pronouns have six-case inflectional paradigms for singular and plural. Noun declension, in addition to case, marks gender (masculine and neuter, and two feminine declensions) and number, with adjectives agreeing for gender, number, and case with the nouns, for example,

\[ \text{bel-aja sten-a} \]

“white Fm Sg wall Fm Sg”

“white wall”
Russian verbs are organized into numerous verb classes varying in degree of regularity, two main conjugational patterns differing by the thematic vowel in the inflections, and two conjugational paradigms (nonpast and past). The nonpast paradigm includes six forms: first, second, and third person singular and plural. Past tense has three forms for masculine, feminine, and neuter, and one for the plural. The system of tenses is very simple: present, past, and future.

Russian has SVO word order as a neutral default setting; however, word order is flexible and primarily reflects topic-comment structure, with the theme introduced at the beginning of the sentence and the rhyme at the end. Questions do not require any verb fronting.

Data collection and analyses were carried out completely in parallel for the two target languages. The same aptitude test in Russian was given to the two groups, and the same kind of grammaticality judgments test was used for both, except that it was necessary, of course, to develop different test items to measure the specific learning outcome in the two languages. Results for the two groups are therefore presented separately.

STUDY 1: RUSSIAN IMMIGRANTS ACQUIRING ENGLISH IN NORTH AMERICA

Method

Participants. The participants in this study were 76 Russian-speaking immigrants above the age of 18, who had acquired English as a second language (ESL). They were living in Chicago, New York, or Toronto. The minimum length of residence in North America was set at 8 years to make sure that ultimate attainment levels had been reached. This is a conservative cutoff point, given that no age effect studies on the acquisition of morphosyntax have reported length of residence effects beyond even the first 5 years; it is also higher than in most age effect studies, equaled only by Flege et al. (1999) and surpassed only by Abrahamsson and Hyltenstam (2008), Ball (1996), Birdsong and Molis (2001), and DeKeyser (2000). The age of acquisition varied from 5 to 71 (see Table 1).

Table 1

These immigrants varied widely in educational background, but the vast majority had college degrees and white-collar jobs; a few even had graduate degrees. Some had attained varying levels of proficiency in one or more languages (Ukrainian, Polish, Georgian, Tajik, Uzbek, Armenian, Romanian, Italian, French, German, Yiddish, or Hebrew) before emigrating; a few had started learning one or two languages (Polish, German, Spanish, Chinese, Japanese, Latin, or Hebrew) after immigrating into North America and learning English. One had lived in Israel and spoken Hebrew for 15 years before moving on to Canada. None, however, had had substantial English teaching or substantial experience using any Germanic or Romance language before emigrating from the (former) Soviet Union.
Table 1. Descriptive statistics for the participants in North America (n = 76)

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOA</td>
<td>5</td>
<td>71</td>
<td>32.54</td>
<td>18.01</td>
</tr>
<tr>
<td>LOR</td>
<td>8</td>
<td>28</td>
<td>11.71</td>
<td>4.03</td>
</tr>
<tr>
<td>AAT</td>
<td>19</td>
<td>76</td>
<td>43.93</td>
<td>17.74</td>
</tr>
<tr>
<td>GJT score</td>
<td>104</td>
<td>198</td>
<td>150.76</td>
<td>27.32</td>
</tr>
<tr>
<td>APT score</td>
<td>5</td>
<td>36</td>
<td>22.58</td>
<td>7.39</td>
</tr>
</tbody>
</table>

Note: AOA, age of acquisition; LOR, length of residence; AAT, Age at testing; GJT, Grammaticality Judgment Test; APT, Aptitude Test.

Instruments

Grammaticality Judgment Test (GJT). A 204-item test was administered to all participants to assess their proficiency in ESL. This instrument was an adapted and shortened version of Johnson and Newport’s (1989) test, largely similar to the DeKeyser’s (2000) adaptation, but with a few extra items to ensure a better representation of the definite article, a category absent in Russian (Chesterman, 1991; Lyons, 1999, Wexler, 1996). The first four items were training items not counted in the analysis.

Aptitude test. Participants’ aptitude was assessed by means of verbal sections of the Russian version of the Inter-University Psychometric Entrance Test (National Institute for Testing and Evaluation, 2001). This version of the test was designed for Russian-speaking college applicants in Israel, and is comparable to the verbal Scholastic Aptitude Test in the United States. This instrument was chosen because it fulfilled the four major requirements of being (a) a test of aptitude, (b) in the participants’ native language, (c) at the right level of difficulty, and (d) usable for both parts of our study in Israel and in North America (no cultural bias was detected in the content of any of the items). The two parts of the test used in this study were sections 3 and 5 (KR-20 reliability = 0.76 for section 3 and 0.71 for section 5, 0.85 for the total of the two), each consisting of 19 multiple-choice items (testing definitions, analogies, and verbal reasoning). For the purpose of this study, aptitude means verbal aptitude in the way it is usually understood in educational psychology, a broader construct than the “modern language learning aptitude” that most SLA research on aptitude focuses on and that is measured by tests such as the MLAT or PLAB (granted, of course, that there are strong correlations between L1 proficiency, verbal aptitude/intelligence, foreign language aptitude, and SLA, in ways that are still poorly understood, but in all likelihood because certain aspects of L1 proficiency and foreign language aptitude are a function of verbal aptitude in the broader sense; see, e.g., Hulstijn and Bossers, 1992; Humes-Bartlo, 1989; Skehan, 1986, 1990; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2006).

As is the case in all verbal aptitude tests, knowledge of the language tested is a factor in the test used here, but there is evidence that this factor played no more role in this Russian version of the test than in the original test: confirmatory factor
analysis and multidimensional scaling have shown the dimensional structures of tests to be equivalent across the Russian and the original Hebrew versions (Allalouf, Bastari, Sireci, & Hambleton, 1997), and where there is differential item functioning for the two versions, it is clearly because of problems of translation/adaptation for specific items and not for lack of construct validity for the test as a whole (Allalouf, Hambleton, & Sireci, 1999). In other words, the test measures verbal aptitude rather than knowledge of language in general or vocabulary specifically, not surprisingly, given that it was designed specifically to test verbal aptitude in the broad sense (as part of a college entrance examination), and moreover for specifically the kind of population we are working with in this study: immigrants who arrived at different ages, and who are bilinguals with somewhat varying levels of L1 Russian.

**Biographical questionnaire.** All participants filled out a three-page questionnaire about their language background, educational background, age, age of arrival, age of acquisition (usually the same as age of arrival, but later in the few cases where participants were not required to use English for communication immediately upon arrival), and current proficiency in English and Russian.

**Procedures**

Participants were recruited via flyers posted in public places, ads in publications aimed at Russian immigrants, and word of mouth. They were paid US $20 or Canadian $30 for participation in the study. They were tested individually or in small groups, in a quiet room, usually at home. After signing the consent form, they filled out the background questionnaire first, then took the grammar test, and finally the aptitude test.

The items on the grammar test were presented auditorily by playing a digitized recording of all sentences, each read twice in a row, with a 3-s interval between the two readings and a 6-s interval between sentence pairs. The sentences were recorded by a female native speaker of English, an ESL teacher, and amateur singer with a very clear voice, in a fixed random order. The entire test took about an hour; there was a 5-min break halfway.

The aptitude test was written; participants could work at their own pace, except that there was a time limit of 25 min for each section.

**Results**

The scores on the GJT ranged from 104 to 198 out of 200, with a mean of 150.76. The reliability coefficient (KR-20) was 0.97.

The relationship between age of age of acquisition and ultimate attainment is represented in Figure 1. The corresponding correlation coefficient is $-0.80 (p < .001)$. This is in line with the correlation coefficients found in other studies (e.g., $-0.77$ in Birdsong & Molis, 2001; $-0.63$ in DeKeyser, 2000; $-0.77$ in Johnson & Newport, 1989), but it does not mean anything in itself; it could hide crucial differences in correlation for various age ranges, as argued in the previous section,
Figures 2, 3, and 4 show the age of acquisition–ultimate attainment relationship for three different age ranges: <18 (n = 20), 18–40 (n = 26), >40 (n = 30), and 4. The scale for the Y axis has been kept constant for all three figures, for ease of comparison. As can be seen in the figures, the regression line is much steeper in Figure 2 (age of acquisition <18) than in Figure 3 (age of acquisition = 18–40), or Figure 4 (age of acquisition > 40). The corresponding correlation coefficients are −.69 (p < .01) for age of acquisition < 18; −.44 (p < .05) for age of acquisition = 18–40; and −.27 (ns) for age of acquisition > 40. More important, however, are the correlations when the effect of age at testing is partialed out (given that, even though the correlation between age of acquisition and age at testing is much smaller in the subsamples than in the total sample, but still not negligible,. 41 for the <18 group, .88 for the 18–40 group, and .83 for the <40 group). When age at testing is partialed out, the difference between the three groups becomes dramatic, because the correlation for the youngest group increases slightly, whereas the

Figure 1. A scatterplot for all ages in North America.
Table 2. Correlations between the main variables for the participants in North America (n = 76)

<table>
<thead>
<tr>
<th></th>
<th>AOA</th>
<th>GJT</th>
<th>AAT</th>
<th>LOR</th>
<th>APT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOA</td>
<td>1</td>
<td>-0.80</td>
<td>0.97</td>
<td>-0.05</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(.00)</td>
<td>(.00)</td>
<td>(.67)</td>
<td>(.02)</td>
<td>(.84)</td>
</tr>
<tr>
<td>GJT score</td>
<td>1</td>
<td>-0.78</td>
<td>0.07</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.00)</td>
<td>(.56)</td>
<td>(.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT</td>
<td>1</td>
<td>0.16</td>
<td>-0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.17)</td>
<td>(.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOR</td>
<td>1</td>
<td>-0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APT score</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: AOA, age of acquisition; GJT, Grammaticity Judgment Test; AAT, age at testing; LOR, length of residence; APT, Aptitude Test. The values in parentheses are p values.

Figure 2. A scatterplot for the age of acquisition at an age of <18 in North America.

correlations for the other two groups are no longer significantly different from zero: \( r = -0.71 \) \((p < .01)\) for age of acquisition <18; \(-0.17 \) (ns) for age of acquisition 18–40; and \(-.12 \) (ns) for age of acquisition >40. (The reverse partial correlation, between age of acquisition and age at testing with ultimate attainment partialed
The role played by aptitude also differs by age group. For all participants combined, the correlation between ultimate attainment and aptitude is .21 (ns); for an age of acquisition of <18 it is .11 (ns), for an age of acquisition of 18–40 it is .44 (p < .05), and for an age of acquisition of >40 it is .33 (ns).

As it is often assumed that the steepest decline in learning takes places around age 12, we did a further analysis splitting the <18 group into a ≤12 group and a >12 group. Within each of these groups, the correlation between the age of acquisition and ultimate attainment is quite small: for age ≤12, it is −.26 (ns, n = 11), and for age >12 it is .01 (ns, n = 12). The correlation coefficients may not be very reliable with such small sample sizes, but the difference between the two groups for the score on the GJT looms large: for the ≤12 group the mean is 187.27; for the >12 group it is 166.42; t (21) = 3.30, p < .01. Thus, it appears that the biggest decline does take place at around age 12.

**Discussion**

**Hypothesis 1.** The hypothesis that the slope of the age of acquisition–ultimate attainment function would not be linear throughout the life span, but instead show a marked flattening between adolescence and adulthood, was confirmed. Even
the raw correlations for the age of acquisition ranges 18–40 and 40+ were flatter than for the 0–17 range; and when the effect of age at testing was partialed out, the effect became quite dramatic because the age of acquisition–ultimate attainment correlation for the age of acquisition <18 group increased slightly to −.71 (p < .01), whereas the correlations for later age ranges of comparable size became very small and nonsignificant. This finding is what one would expect under the CPH: after this period is over, one no longer expects to see the same decline (even though some decline for other reasons is expected, of course, especially for the oldest participants). A further analysis shows the decline to be especially steep around age 12 (with the caveat that the sample sizes for age ≤ 12 and age = 13–18 are quite small).

Hypothesis 2. The hypothesis that the relationship between aptitude and ultimate attainment would differ markedly for the younger and the older arrivals was also confirmed. The correlation for the age of acquisition <18 group was very small (r = .11) and nonsignificant; for the age of acquisition 18–40 range it was substantial and significant (r = .44; p < .05). For the oldest arrivals, whose age of acquisition was over 40 and whose age at testing varied from 50 to 76, with a mean of 63, the correlation flattens somewhat again (r = .33, ns), presumably because other factors were playing an increasing role in determining test performance for this age range.
Table 3. Descriptive statistics for the participants in Israel (n = 62)

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOA</td>
<td>4.12</td>
<td>65.2</td>
<td>30.57</td>
<td>16.94</td>
</tr>
<tr>
<td>LOR</td>
<td>8.46</td>
<td>28.86</td>
<td>12.27</td>
<td>3.59</td>
</tr>
<tr>
<td>AAT</td>
<td>16.25</td>
<td>75.17</td>
<td>42.84</td>
<td>16.55</td>
</tr>
<tr>
<td>GJT score</td>
<td>101</td>
<td>196</td>
<td>149.58</td>
<td>26.33</td>
</tr>
<tr>
<td>APT score</td>
<td>1</td>
<td>36</td>
<td>19.84</td>
<td>8.59</td>
</tr>
</tbody>
</table>

Notes: AOA, age of acquisition; LOR, length of residence; AAT, age at testing; GJT, Grammaticality Judgment Test; APT, Aptitude Test.

STUDY 2: RUSSIAN IMMIGRANTS ACQUIRING HEBREW IN ISRAEL

Method

Participants. The participants in this study were 62 Russian-speaking immigrants above the age of 18, who had acquired Hebrew as an L2. All lived in communities close to Tel-Aviv. The minimum length of residence in Israel was set at 8 years to make sure that ultimate attainment levels had been reached. The age of acquisition varied from 4 to 65 (see Table 3). The immigrants varied in educational background. Most people in the sample had 13–18 years of schooling. Academic degrees were mostly in science, one-quarter had degrees in humanities and sociology, and the rest had degrees in the life sciences. Two-thirds had graduated from a Russian-speaking university, compared to one-third who had received their academic degree in Israel. Most participants were working or had worked in Israel by the time of the study. A scale of high (e.g., engineer, physician), middle (e.g., teacher), and low positions (e.g. cashier, cleaner) was constructed to evaluate immigrants’ work positions. It showed that most participants had high or intermediate positions.

Most participants knew other languages, in addition to Russian and Hebrew, before emigrating from the (former) Soviet Union: Ukrainian, Polish, Georgian, Romanian, French, German, Yiddish, or English. Most participants did not know any Hebrew before emigrating; those who did have knowledge had mostly “poor” knowledge of the language. Most of them had studied Hebrew in Ulpan (i.e., intensive immersion Hebrew language classes for new immigrants, provided by the state, which finances immigrants’ living expenses during their first months in Israel, so they can devote more time to language learning) and had also taken a course or studied at a Hebrew-speaking institute (school, college, university).

Instruments

GJT. All participants took a GJT in Hebrew consisting of 204 items representing six basic categories of Hebrew morphology, such as noun–adjective agreement, use of the definite article, and morphology of past, present, and future tense. The test was designed by the second and third authors, specifically for the purpose of this study (see Appendix A for a list of structures and examples).
Aptitude test. The same aptitude test was used as in the North American study.

Biographical questionnaire. All participants filled out an extensive biographical questionnaire of 66 multiple-choice and open-ended questions about their age of acquisition, age at testing, length of residence, gender, academic background, profession, children born in Israel, economic situation, self-assessment of Hebrew knowledge at the time of testing, self-assessment of Hebrew knowledge prior to immigration, sources of Hebrew knowledge, contexts of Hebrew usage, language preferences, identity, and motivation.

Procedures

Participants were recruited via flyers posted in public places, ads in publications aimed at Russian immigrants, and word of mouth. They were paid US $20 for participation in the study. They were tested individually, in a quiet room, usually at home. After signing the consent form, they filled out the background questionnaire, then took the grammar test, and finally the aptitude test.

The items on the grammar test were presented auditorily by playing a digitized recording of all sentences, each read twice in a row, with a 3-s interval between the two readings and a 6-s interval between sentence pairs. The sentences were recorded by the second author, a linguist and proficient speaker of Hebrew, in a fixed random order. The entire test took about 1 hr; there was a 5-min break halfway through.

The aptitude test was written; participants could work at their own pace, except that there was a time limit of 25 min for each section.

Results

The scores on the GJT ranged from 101 to 196 out of 204, with a mean of 150. The reliability coefficient (KR-20) was 0.98.

The relationship between age of acquisition and ultimate attainment is represented in Figure 5. The corresponding correlation coefficient is \(-0.79 (p < 0.001)\). As Table 4 shows, however, ultimate attainment is also strongly correlated with age at testing, but not significantly correlated with length of residence. Therefore, length of residence was not included in subsequent analyses, but age at testing was used as a control variable.

Figures 6, 7, and 8 present scatterplots of the age of acquisition–ultimate attainment relationship for three different age ranges: <18 (n = 17), 18–40 (n = 32), >40 (n = 13). The scale for the Y-axis has been kept constant for all three figures, in order to facilitate comparisons. The corresponding correlation coefficients are \(-0.48 (p = 0.05)\) for age of acquisition <18; \(-0.37 (p < 0.05)\) for age of acquisition = 18–40; and \(-0.53 (ns)\) for age of acquisition > 40.

In accordance with the North American data, it is important to look at the correlations when the effect of age at testing is partialed out (given that, even though the correlation between age of acquisition and age at testing is much smaller in the subsamples than in the total sample, but still not negligible at .79 for the <18 group, .88 for the 18–40 group, and .98 for the <40 group). When
the effect of age at testing is partialed out, the difference between the youngest group and the two older groups becomes quite large, because the correlation for the youngest group remains moderate and significant (−.51, \( p < .05 \)), whereas the correlations for the other two groups are small and not significantly different from zero: \(-.12\) (ns) for age of acquisition = 18–40; and \(-.33\) (ns) for age of acquisition > 40. (The reverse partial correlation, between age of acquisition and age at testing with ultimate attainment partialed out, is never significant: .29 for the <18 group, -.08 for the 18–40 group, and .23 for the >40 group.)

The role played by aptitude also differs by age group. For all participants together the correlation between ultimate attainment and aptitude is −.003 (ns); for age of acquisition < 18 it is −.37 (ns); for age of acquisition = 18–40, \( r = .45 \) (\( p < .01 \)); for age of acquisition > 40, \( r = .14 \) (ns).

As was done for the North American group, we conducted a further analysis splitting the <18 group into a ≤12 group and a >12 group. Within each of these groups the correlation between age of acquisition and ultimate attainment is quite small: for age ≤12 it is −.38 (ns, \( n = 13 \)); for age >12 it is .008 (ns, \( n = 7 \)). Again, the correlation coefficients may not be very reliable with such small sample sizes, but the difference between the two groups for the score on the GJT looms large here as well: for the ≤12 group the mean is 181.7; for the >12 group it is 158.7;
Table 4. Correlations between the main variables for the participants in Israel (n = 62)

<table>
<thead>
<tr>
<th></th>
<th>AOA</th>
<th>GJT</th>
<th>AAT</th>
<th>LOR</th>
<th>APT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOA</td>
<td>1</td>
<td>-.79</td>
<td>.98</td>
<td>-.21</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.00)</td>
<td>(.00)</td>
<td>(.10)</td>
<td>(.19)</td>
</tr>
<tr>
<td>GJT score</td>
<td>1</td>
<td>-.77</td>
<td>.19</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.00)</td>
<td>(.08)</td>
<td>(.98)</td>
<td></td>
</tr>
<tr>
<td>AAT</td>
<td>1</td>
<td>-.001</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.99)</td>
<td>(.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOR</td>
<td>1</td>
<td>-.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APT score</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: AOA, age of acquisition; GJT, Grammaticality Judgment Test; AAT, age at testing; LOR, length of residence; ATP, Aptitude Test. The values in parentheses are p values.

\[t(16) = 2.37; p < .05.\] Thus, it appears that the biggest decline does take place at around age 12 for the group in Israel.

Discussion

Hypothesis 1. The hypothesis that the slope of the age of acquisition–ultimate attainment function would not be linear throughout the life span, but instead show a marked flattening between adolescence and adulthood, was confirmed. Even the raw correlation for the age of acquisition range 18–40 was much flatter than for the 0–17 range, and when the effect of age at testing was partialed out, the effect became quite dramatic, in the sense that the age of acquisition–ultimate attainment correlation for the <18 group was moderate \((r = -.48)\) and significant, whereas the correlations for later age ranges of comparable size became very small and nonsignificant. This finding is what one would expect under the CPH: after this period is over, one no longer expects to see the same decline. Some decline for other reasons is expected, of course, especially for the oldest participants; it is found here quite clearly for the oldest participants (age of acquisition > 40; age at testing = 50.2–75.0 with a mean of 67.8), but it disappears completely when age at testing is partialed out. A further analysis shows the decline to be especially steep around age 12 (with the caveat that the sample sizes for ages of ≤12 and 13–18 are quite small).

Hypothesis 2. The hypothesis that the relationship between aptitude and ultimate attainment would differ markedly for the younger and the older arrivals was also confirmed. The correlation for the age of acquisition <18 group was small \((r = -.37)\) and nonsignificant; for the age of acquisition 18–40 range it was substantial and significant \((r = .45; p < .01)\). For the oldest arrivals, whose age of acquisition is over 40 and whose age at testing varies from 50.2 to 75.2 with a
mean of 67.8, the correlation flattens again ($r = .17, ns$), presumably because other factors are increasingly playing a role in determining test performance for this age range.

**GENERAL DISCUSSION**

The results for the samples from North America and Israel show remarkably similar patterns, despite the radical differences in L2 structures to be acquired (morphology-rich Hebrew vs. morphology-poor English) and the different societal context. All the learners had in common in both cases was their native language (besides, for many of them, ethnic and religious affiliation, and perhaps attitudes toward language and schooling).

For younger learners (below the age of 18), ultimate attainment in grammar was strongly predicted by age of arrival, but not by aptitude. For young adults (ages 18 to 40), it was the other way around: aptitude, but not age of arrival predicted the level of ultimate attainment. For the oldest learners, who were over age 40 on arrival and typically between 50 and 75 at testing, neither aptitude nor age of arrival were good predictors, only age at testing.

The findings about the effect of age of arrival are perfectly compatible with the predictions of the CPH: a rather precipitous decline in the ability to acquire a
language during a time period ending somewhere in adolescence, followed by a period of no further decline as a function of age of arrival (even though there may be some decline because of other factors, such as age at testing, especially for the oldest participants). These findings concur with those of studies that have shown a pattern of rapid decline followed by relative stability (e.g., DeKeyser, 2000; Flege et al., 1999; Johnson & Newport, 1989). They are different from those found in studies such as Hakuta et al. (2003), who found a decline throughout the life span, and they are the opposite of those in Birdsong and Molis (2001), the only study in the literature that found no decline for the younger group, but a significant decline for the adults. Elsewhere (DeKeyser, 2006) we provided explanations for why the latter studies may have found such lack of stabilization in adulthood: for example, measurement of ultimate attainment with nothing but a very coarse-grained self-assessment in the case of Hakuta et al. (2003) and the presence of some outliers in Birdsong and Molis (2001). The present study suggests that not taking into account age at testing, usually substantially confounded with age of acquisition, may have been another important reason for their findings and their discrepancy with ours. In contrast, the lack of decline in the early learner group in Birdsong and Molis may have been because of the L1–L2 combination: when the two languages are relatively closely related such as English and Spanish, one would expect the decline in the early group (because of the critical period) to be

Figure 7. A scatterplot for the age of acquisition at an age of 18–40 in Israel.
Figure 8. A scatterplot for the age of acquisition at an age of >40 in Israel.

less marked, and therefore, it would look more similar to the decline in the late group (due largely to confounding with age at testing). In our two studies here, both L1–L2 combinations (Russian–English and Russian–Hebrew) were more challenging.

The findings on the effect of aptitude are also compatible with previous findings, in this case on the role of aptitude at various ages (DeKeyser, 2000; Harley & Hart, 1997), despite the very different operationalization of aptitude in those studies (tests of aptitude for foreign language learning) compared to the present one (a broader verbal aptitude test).

Together these findings provide both evidence for a quantitative decline of learning ability and a qualitative shift in grammar learning mechanisms as a function of age before adulthood; they contradict the claims that there is no quantitative evidence of a critical period because there is no discontinuity in the decline (e.g., Birdsong, 2004, 2005, 2006; Hakuta et al., 2003) or because there is no evidence of qualitative differences as a function of age (Hakuta, 2001). Putting both age at testing and aptitude into the picture has provided a dramatically different picture for younger compared to older learners of how much learning takes place and how; younger learners learn more while relying less on aptitude; older learners learn less, and to the extent they do learn, must rely more heavily on their verbal aptitude. These findings should not be immediately generalized to
all aspects of L2, of course; only morphosyntax was studied here, not the lexicon, pragmatics, or pronunciation.

We are now conducting a fine-grained study of the linguistic aspects of the various morphosyntactic structures in our tests to shed further light on the nature of the qualitative differences in learning mechanisms for children and adults. However, we do not want to suggest, of course, that age of arrival and aptitude are the only variables that matter in determining ultimate attainment. A wide variety of studies have documented a very large spread in proficiency among adult learners, due not only to age and aptitude, but also to personality, motivation, and level of education, among other variables (see, e.g., Dörnyei, 2005; Dörnyei & Skehan, 2003; Ellis, 2004). These variables do not take away from the importance of the age factor, however; on the contrary; studies that have investigated level of education and age of arrival in the same data set have found that, although level of education is a predictor of ultimate achievement, the shape of the age of arrival–ultimate attainment function is the same for learners with different levels of education (Bialystok & Hakuta, 1999; Hakuta et al., 2003). At this point it would be very premature to discount the importance of age of arrival as an independent predictor of ultimate achievement in L2 grammar. Future studies will need to take age at testing into account when analyzing the relationship between age of acquisition and ultimate attainment. Another important improvement over existing research, our own work included, would come from still larger numbers of participants, but without sacrificing the quality of the data. Conversely, what is perhaps most needed in this area of research at this point is the use of a wider variety of fine-grained dependent measures, not just grammaticality judgments or global accuracy ratings. It is also desirable, everything else being the same, to have a population of immigrants who are strictly monolingual at the time of migration and a precise documentation of the amount and quality of L1 and L2 use by these immigrants after the onset of acquisition.

APPENDIX A

**Structures of the Hebrew GJT (100 item pairs)**

1. Noun plurals \((N = 10)\), for example, *Dani kana harbe maxshir be-hom senter lifney shavua/Dani kana harbe maxshir-im be-hom senter lifney shavua* “Danny bought a lot of tool at Home Center last week/Danny bought a lot of tool, Pl at Home Center last week.”
2. Adjective inflection \((N = 32)\), for example, *Ron kibel shanox le-yom ha-huleket shelo/Ron kibel shan ox le-yom ha-huleket shelo* “*Ron has received a black, Fem watch for him birthday/Ron has received a black watch for him birthday.”
3. Verb inflection \((N = 8)\), for example, *lama at medabet ir ha-tipus ha-ze bixal?/lama at medaber ir ha-tipus ha-ze bixal?* “*Why are you talking to this creature anyway?/Why are you talking, Fm to this creature anyway?*
4. Morphosyntactic constructions, for example, compounding, subordination, conditionals \((N = 16)\) *im Dan yecaxceax shinayim hayu lo shinayim nekiyotim Dan yecaxceax shinayim yihyu lo shinayim nekiyot* “*If Dan brushes his teeth he had clean teeth/If Dan brushes his teeth he will have clean teeth.*”
5. The definite article \((N = 26)\), for example, *tavi li bevakasha magevet me-aron tavi li bevakasha magevet me-ha-aron* “*Please bring me a towel from closet/Please bring me a towel from the-closet.*”
DeKeyser et al.: Cross-linguistic evidence for the nature of age effects in SLA

Numeral agreement ($N = 8$), for example, *pagashnu shalosh banim ba-xufsha shelanu be-eilat*/*pagashnu shlosha banim ba-xufsha shelanu be-eilat* “We have met three, Fem boys at our holiday in Eilat/We have met three boys at our holiday in Eilat.”

ACKNOWLEDGMENTS
This research was funded by NIH (NICHD) Grant 1 R03 HD41479–01. The authors gratefully acknowledge the help of Becky Bird and Neta Abugov with data collection and Elaine Rubinstein and Gabi Lieberman with data analysis. Parts of this study were previously presented at the annual meeting of the American Association of Applied Linguistics in Portland, Oregon, and at the International Symposium on Bilingualism in Barcelona, Spain.

NOTES
1. A variety of research projects at the Center for the Advanced Study of Language (College Park, MD) are aimed at designing a battery of aptitude tests with better validity, especially for advanced stages of language learning.
2. Eight more people were tested, but the data for seven of them were not entered into the analysis because the questionnaire showed that they did not meet the criteria for the study (they fell slightly short of 8 years of residence or had a slight hearing problem), and one person’s data were eliminated from the analysis because he had a score below chance on the GJT, presumably because of ignoring or misunderstanding the instructions.
3. Eleven more people were tested, but their data were eliminated from the analysis because they had a GJT score below chance, presumably because of ignoring or misunderstanding the instructions, or scored zero on the aptitude test.
4. Two participants whose age of acquisition was $<3$ were eliminated from the analysis after the comment from an external reviewer that one cannot speak of SLA at such a young age. The minimum is now set at age 4.
5. Two people scored zero on the aptitude test, presumably because they misunderstood or ignored the instructions; they were eliminated from the analyses involving the aptitude test.

REFERENCES
Berman, R. A. 1999. Relevant features of spoken and written English. In R. Aisenman (Ed.), Working papers in “developing literacy across genres, modalities, and languages” (pp. 4–19). Tel Aviv: Tel Aviv University.


DeKeyser et al.: Cross-linguistic evidence for the nature of age effects in SLA


