Language Learning & Language Teaching

Individual Differences and Instructed Language Learning

EDITED BY
Peter Robinson

OFFPRINT

This is an offprint from:

Peter Robinson (ed.)
Individual Differences and Instructed Language Learning.
John Benjamins Publishing Company
Amsterdam/Philadelphia
2002
(Published as Vol. 2 of the series
LANGUAGE LEARNING & LANGUAGE TEACHING,
ISSN 1569-9471)

ISBN 90 272 1693 2 (Hb; Eur.) / 1 58811 230 6 (Hb; US)
ISBN 90 272 1694 0 (Hb; Eur.) / 1 58811 231 4 (Pb; US)

CHAPTER 11

Aptitude-exposure interaction effects on Wh-movement violation detection by pre-and-post-critical period Japanese bilinguals*

Steven Ross, Naoko Yoshinaga, Miyuki Sasaki

Kwansei Gakuin University / Hirosaki Gakuin University / Nagoya Gakuin University

Introduction

Research on apparently exceptional adult second language acquirers has indicated that there may be factors other than transfer and age of acquisition affecting acquisition of both peripheral and core aspects of a second language grammar. The influence of such individual difference factors has not been extensively examined against a second language acquisition (SLA) theory using the critical period as a developmental benchmark. This study examines Japanese learners' intuition about violations of *that*-trace effect and the subjacency principle. Subjects represent three different samples of critical period-sensitive language learning experience profiles: a) extensive naturalistic exposure to English in childhood; b)teenage SLA through immersion; and c) foreign language learning exclusively. Measures of *that*-trace and subjacency violation grammaticality judgements are regressed on language learning aptitude and pre-and-post critical period language acquisition experience profiles in order to test the hypothesis that individual difference factors interact with pre-and-post-critical period language acquisition experience. Results suggest that individual differences in aptitude interact in subtle ways with learning experience to influence post-critical period SLA.

Theoretical background

Recent work in second language acquisition has begun to explore variation in the post-critical period acquisition of near native-like competence in both peripheral and core grammar domains. Following early claims of complete inaccessibility to Universal Grammar for post-critical period second language learners (Birdsong, 1999; Bley-Vroman, 1989; Schachter, 1989, 1990) recent research suggests that there may be factors which may serve to compensate for post-critical period learners' diminished access to language specific aspects of second language morphosyntax (DeKeyser, 2000; Harley & Hart, 1997, this volume; Newport, 1990) and possibly even Universal Grammar-dependent phenomena (Ioup, Houstagui, El Tigi, & Moselle, 1994; White and Genesee, 1996). Exceptional language learner characteristics (Novoa, Fein & Obler, 1988; Smith & Tsimpli, 1995) are potentially akin to the kinds of individual differences phenomena Bley-Vroman (1990) hypothesized to motivate adult 'problem solving' processes that come into play in post-critical period SLA.

Just as extensive exposure to any natural language during childhood leads to virtual homogeneity in native competence, exposure to a second language in the post-critical period results in a veritable bell curve. Some complex of social, psychological and cognitive processes apparently serves to create variance in proficiency among adults (Skehan, 1989, 1998a, this volume) to such a degree that the usual picture of adult second language learning is one more obvious for its shortcomings than for its successes.

Recent research has begun to explore cognitive factors that may generate the variance in language acquisition outcomes for adults. How deeply individual difference factors can go to account for apparent anomalous cases of post-critical period SLA have yet, however, to be identified.

Aptitude research

Second language learning aptitude (henceforth, aptitude) is traditionally defined as the ability to succeed in learning a foreign language given adequate instruction and/or experience. Ever since Carroll and Sapon's (1959) pioneering development of the Modern Language Aptitude Test (MLAT), research related to aptitude has continued to support four hypotheses: (1) aptitude is typically stable, and unsusceptible to short-term training (e.g., Carroll, 1981; Politzer & Weiss, 1969; Skehan, 1989, but see also Nayak, Hansen, Krueger, & McLaughlin, 1990; McLaughlin, 1990, 1995, reporting that multilingual expe-

riences may enhance the learners' readiness to learn a new language); (2) aptitude consists of several mutually distinct cognitive abilities (e.g., phonemic coding ability, grammatical sensitivity, rote learning ability for foreign language materials, and inductive language learning ability proposed by Carroll, 1981, 1985, 1991; see also the three factor model proposed by Skehan, 1998a, this volume); (3) aptitude is at least partially independent of other cognitive abilities such as general intelligence (e.g., Gardner & Lambert, 1965, 1972; Ganschow, Sparks, & Javorsky, 1998; Robinson, this volume; Sasaki, 1996; Sternberg, this volume; Wesche, Edwards, & Wells, 1982), and (4) aptitude generally has higher and more consistent correlations with second language (L2) proficiency, whether it is acquired formally or informally, than have other individual difference variables such as cognitive styles and personality (e.g., Ehrman & Oxford, 1995; Reves, 1982; Skehan, 1989, 1998a).

The relationship between L2 proficiency and aptitude was first investigated through validation studies of existing aptitude batteries (Carroll, 1958, 1979, 1981; Carroll & Sapon, 1959; Pimsleur, 1966b). Aptitude batteries such as the MLAT and the Pimsleur Language Aptitude Battery (PLAB) correlated moderately (generally between $r = .2$ and .6) with subsequent proficiency tests and final foreign language grades. With homogeneous sample populations, the correlation sometimes went up as high as $r = .7$ (Carroll & Sapon, 1959). However, such validation studies were typically conducted in the psychometric/structuralist tradition, and "researchers in this area have not been excessively influenced by other developments in second language acquisition and learning" (Skehan, 1998a, p. 186).

Skehan's studies in the 80's were among the first to introduce acquisition and learning perspectives into aptitude research. For example, Skehan (1980) focused on memory, one aspect of multidimensional aptitude, and looked at the correlation between the test scores of various memory tests and other measures of English-speaking adults' success in learning Arabic. Skehan reported that tests for "response integration" (memory for unknown language structures) and "memory for text" predicted the participants' language learning success better than those for other types of memory. Subsequently, Skehan (1986a) analyzed the same data, and found two different types of successful language learners: One was the relatively young group that depended more on their memory, and the other was the relatively older group that depended more on their analytic abilities (see also Wesche, 1981 for similar findings).

Later in the 80's, Skehan (1986a, 1986b, 1989) further investigated "the origin of aptitude" (1989, p.194) issue by examining the relationship among

three latent traits: first language (L1) development, aptitude, and foreign language achievement. English speaking children's L1 development was measured by such indicators as mean length of utterance and sentence structure complexity when they were three to four years old. Later, when they entered secondary school (ages 13 to 14), their aptitude and foreign language success were measured. Skehan (1989) reported significantly high positive correlations (above $r = .4$) between some of the L1 indices and those of later aptitude. In fact, these figures were generally higher than the correlations between first language measures and foreign language achievement measures. Based on these results, Skehan posited a latent trait of "pre-programmed autonomous language learning ability" (Skehan, 1989, p. 33) underlying both first language development and second language learning aptitude.

In their analysis of individual differences in French proficiency, Harley and Hart (1997, this volume) identified two patterns of language learning aptitude that correlated differently with proficiency outcomes. Harley and Hart compared the relationship between three different measures of aptitude (associative memory, memory for text, and analytical ability) and various L2 proficiency measures among learners starting to learn French before and after adolescence (early and late immersion students respectively). Harley and Hart first found that the early immersion students' scores on three aptitude subtests were not significantly higher than those of the late immersion students, which contradicts the findings of some previous research (e.g., McLaughlin, 1990) that aptitude can be developed through language learning experiences. Furthermore, Harley and Hart reported that early immersion students' L2 learning success tended to be associated with the memory measures (especially the memory for text measure) whereas late immersion students' L2 learning success was associated with the analytical ability. Of interest was the distinction between pre-critical period learners' 'verbal memory' aptitude scores and post-critical period learners' 'language analysis' aptitude scores, which were correlated with several measures of French proficiency development. These differences between the two profiles of learners are of particular relevance to the present study because they suggest the possibility of an aptitude by experience interaction.

The potential influence of experience interacting with aptitude is also seen in research undertaken by DeKeyser (2000), who found evidence corroborating the Harley and Hart observation that language learning aptitude interacts with initial exposure to a second language. DeKeyser, in replicating Johnson and Newport's (1989, 1990) studies using grammaticality judgments, observed

that adults' ability to perform 'language analysis' correlates with English proficiency for Hungarian-Americans, whose exposure to English began in the post-critical period. The correlation between language analysis and proficiency was seen to diminish into insignificance for those subjects whose acquisition began early, presumably before the critical period, relative to the robust correlation between language analysis and proficiency among adults. For peripheral aspects of English and French grammars at least, these two studies suggest that individual differences in language learning aptitude influence child and adult language learning success in different ways. It appears that child second language acquirers might not need aptitude for success, possibly because of the immutable influence of Universal Grammar. If neurological maturation leads to the eventual atrophy of language learning, individual differences among adults may reflect variation in neurological organization (Eubank & Gregg, 1999; Smith & Tsimpli, 1995), or possibly the emergent influence of other cognitive factors such as language learning aptitude.

To date, few studies have found instances of post-critical period SLA resembling native-like competence (Birdsong, 1992; Coppieters, 1987; Ioup, et al. 1994). When such cases are observed, transfer of first language parameters has been the preferred explanation (see White, 1996 for a review) In examining near-native intuition by adults without a specific consideration of transfer, White and Genesee (1996) interviewed adults who apparently had achieved near native-like performance. The object of their interest was to determine if such adults could demonstrate native-like competence in detecting 'strong' island violations — extraction out of complex noun phrases, relative clauses, adjuncts, and subjects in the process of *wh*-question formation. White and Genesee concluded that indeed there was a consistent pattern of 'no difference' between their near-native subjects and the native controls, and more significantly, that there was no effect for age of acquisition (cf. Eubank & Gregg, 1999).

White and Genesee's study suggests that even post-critical period second language learners "can achieve native-like competence with respect to constraints of UG, provided care is taken to ensure native-like proficiency" (1996, p.251). Of interest to SLA research is how such learners *get* native-like proficiency — given the widespread observation that the vast majority of adults come nowhere close to native-like L2 competence or performance.

It now appears that cognitive accounts of individual differences consistently show effects for formal learning and language-specific aspects of morphosyntax. The extent to which these individual differences interact with pre-

and-post critical period second language acquisition experience profiles sets a new research agenda. In order to examine this interaction, we operationalize grammar analysis aptitude, which we assume to approximate Bley-Vroman's (1990) notion of 'problem solving' ability in adult learners.

Wh-movement phenomena

This section describes the subjacency condition and the empty category principle (ECP), which are considered to be the constraints relevant to the formation of *wh*-questions, following the traditional principles-and-parameter approach[1]. We will use a pre-minimalist version of syntactic theory (i.e., the government-and-binding theory) since it is widely known and is easily accessed in the literature. The main focus is on three constructions, which are of interest in the present study: constructions with subjacency violations involving *wh*-extraction out of relative clauses and out of adjunct clauses, and a construction with an ECP violation involving the *that*-trace effect, which will be clarified in what follows. We briefly discuss these principles in relation to the SLA research on UG accessibility issues.

The subjacency condition and the empty category principle

In English, *wh*-questions are formed with the *wh*-word moved to the sentence initial position regardless of its underlying position, as *what* in *What did John eat?* In the tradition of generative grammar, such *wh*-words are considered to be extracted from their underlining position to the sentence initial position, leaving their trace behind (see e.g., Haegeman, 1994, pp. 371–372).

English *wh*-question formation involving the extraction of *wh*-words out of clauses, shown in (1a) and adjunct clauses shown in (1b) are not permitted, leading to unacceptability of the sentences (see e.g., Chomsky, 1986, p.34 & p.31).

(1) a. *What did Kate like the story that described_?
b. *Who did Mary see the doctor after the rock hit_?

In sentence (1a), *what* originates in the position of the direct object of the verb *described*, which is in the clause modifying the noun *story*; in sentence (1b) *who* originates in the position of the direct object of the verb *hit*, which is in an adjunct clause beginning with *after*.

The subjacency condition rules out those structures. In the *Barriers*

framework (Chomsky, 1986, p.30), this is explained in terms of the notion of barriers; the movement of *wh*-words over more than one barrier leads to unacceptability by native speakers of English. The extraction of the *wh*-word in both structures illustrated in (1) involves more than one barrier, violating the subjacency condition[2].

Wh-questions such as in (2a), where the *wh*-word originates in subject position in the embedded clause, are often considered to be ungrammatical unless the complementizer *that* is deleted. The comparable grammatical sentence without *that* is shown in (2b).

(2) a. *What did Mary think that broke the window?
b. What did Mary think broke the window?

This phenomenon, known as the *that*-trace effect (see e.g., Haegeman, 1994, pp.398–399), is accounted for by the empty category principle (ECP, Chomsky, 1986, pp.47–48). The ECP rules out sentences with empty categories (such as traces of *wh*-words) from being properly governed; the existence of the complementizer '*that*' prevents the trace of *wh*-word in subject position in a sentence like (2a) from being properly governed, hence violating the ECP (e.g., Chomsky, 1986; Lasnik & Saito, 1984).

UG principles and studies of UG access

The analysis of UG principles like the subjacency condition and the ECP in the performance of learners whose native language lacks overt *wh*-movement comparable to English has been an important tool in the issue of the access to UG (see e.g., Bley-Vroman, Felix, & Ioup, 1988; Schachter, 1989, 1990). As Herschensohn (1999) points out, even though linguistic theory has changed in recent years, variance in judgement of subjacency violations has not adequately been explained and thus remains a viable object of study.

If learners of English whose native language lacks syntactic *wh*-movement (such as Korean or Japanese) show knowledge of the subjacency condition in judging English *wh*-question sentences, their knowledge is not directly attributable to transfer from their L1; neither is it likely to be available from L2 input (e.g., Bley-Vroman et al., 1988, p.5; see White, 1989, p.74; 1992, pp.448–49).

The prevailing interpretation is that input-independent and transfer-independent accurate judgements about UG principles indicates continued access to those principles (see e.g., Martohardjono & Gair, 1993; White, 1989). If, on the other hand, learners fail to show the knowledge of such UG princi-

ples, it can be taken to suggest no more access to UG; however, it has also been claimed that inaccurate judgements do not necessarily show inaccessibility to UG principles (White,1989; Martohardjono & Gair, 1993).

White (1992, p.448) warns that it is very difficult to find transfer-independent interlanguage syntax (e.g., subjacency operates in L2 but not in L1), referring to the possibility that learners may make use of L1 knowledge of subjacency which constrains movement of other structures in their L1. If such knowledge is transferable, it should help learners to accurately reject relevant violations.

It has been reported that second language learners are somewhat less accurate than native speakers in rejecting ungrammatical sentences involving *wh*-movement (White, 1992, p.449). In Schachter (1989), for example, many subjects, especially Korean learners of English, did not perform well on detecting subjacency violations (Schachter, 1989, p.85). Bley-Vroman et al.'s results showed that Korean learners of English performed better than chance, but they performed well below native speaker norms (Bley-Vroman et al., 1988, p.27).

Second language acquisition researchers who support the continuity hypothesis (access to UG) attempt to account for the phenomenon that learners do not perform like native speakers by arguing that the learners may analyze sentences differently from native speakers, and that inaccuracy of the learners does not necessarily mean that they do not have access to UG (Martohardjono & Gair, 1993; White, 1992; cf. Yusa, 1999). For example, White (1992) claims that Korean or Japanese learners who accept English sentences with subjacency violations analyze those English sentences as involving no movement, since such an analysis is possible in the learners' L1; thus subjacency is not relevant and they are not violating principles of UG.

As for the ECP, it is considered possible in the principle-and-parameter model (pre-minimalist program model) that *wh*-words move at the level of Logical Form (LF) in languages such as Japanese and Korean which do not have syntactic *wh*-movement. This type of movement in LF is subject to the ECP; that is, the ECP is used in those languages (White, 1989, p.74; Yusa, 1999, p.291). White notes that the ECP is used in Korean, but not for syntactic *wh*-movement. White further notes that learners could detect ECP violations via knowledge of their native language, although Bley-Vroman et al. (1988) did not find strong evidence of it.

White and Genesee (1996, p.245) report that judgements on *that*-trace effect among their pilot native speakers of English were not consistent. They

attributed this fact to the claim that there may be dialectal variations for *that*-trace effects (see Rizzi, 1990, p.53).

Japanese phenomena

Unlike English, Japanese lacks overt *wh*-movement. *Wh*-phrases can remain in their underlying positions. The assumption, however, that Japanese completely lacks syntactic *wh*-movement has been questioned. Japanese may have other types of overt movement subject to the subjacency condition, which may undermine the claim of complete non-transferability from Japanese.

For example, Grewendorf (1996, pp.727–728) cites Watanabe's null operator movement in Japanese and Saito's claim that scrambling out of relative clauses and adjunct clauses is subject to the subjacency condition.

Watanabe (1992) claims that a null *wh*-operator associated with the *wh*-phrase, which is phonetically unrealized, moves at overt syntax (S-structure). His claim is based on the observation that sentences such as (3a), where a direct object *wh*-phrase is in a *wh*-island, are marginally unacceptable. Sentences such as (3a) are therefore sensitive to the subjacency condition (Watanabe, 1992, pp.257–8).

(3) a. (?)John-wa [Mary-ga nani-o katta ka dooka] shiritagatte iru no?
John-Top Mary-Nom what-Acc bought whether know-want Q
'What does John want to know whether Mary bought?'

b. John-wa [nani-o katta hito]-o sagashiteiru no?
John-Top what-Acc bought person-acc looking-for Q
'What is John looking for the person who bought?'
(Watanabe, 1992, p.257; 3a is Watanabe's (5b) and 3b is his (5a))

Watanabe (1992), however, also notes that unlike sentence (3a), which involves a *wh*-island, sentence (3b), where the *wh*-word *nani-o* inside of the relative clause, is grammatical. He follows the pied-piping analysis that Nishigauchi (1986, 1990) proposed to account for this (see also Groat & O'Neil, 1996). Setting aside the technical details here, this pied-piping operation is done at LF, and might not be transferred to incorrectly lead Japanese learners to judge English *wh*-questions involving extraction out of a relative clause as acceptable.

If Watanabe's claim is correct, it implies that syntactic *wh*-movement sensitive to the subjacency condition exists in Japanese. It follows that Japanese learners of English may be able to apply their L1 knowledge to detect the ungram-

maticality of such sentences if they "are at least familiar with the fact that overt *wh*-movement is constrained in some way" (Grewendorf, 1996, p.727).

Another possible source of positive transfer from Japanese discussed in the literature is the fact that Japanese has an overt movement — scrambling. Saito (1992, p.72) notes that the sentence in (4), in which a *wh*-phrase is scrambled out of a relative clause, is a case of subjacency violation[3].

(4) ?*Dono hon -o$_i$ [Hanako-wa [NP t_i kaita hito] -ni atta] no
which book-Acc -Top wrote person-to met Q
'Which book$_i$, [Hanako met the person who wrote t_i]'
(Saito, 1992, p.72, Saito's (6))

According to Saito (1985), scrambling out of adjunct clauses leads to unacceptable sentences, as shown in (5).

(5) ??Sono hon-o$_i$ John-ga [Mary-ga t_i yomioete kara] dekaketa (koto)
that book-acc -nom -nom finish-reading after went-out fact
(John went out after Mary finished reading that book)
(Saito, 1985, p.247, Saito's (147a))

Thus if these analyses are correct, it can be predicted that Japanese learners of English may correctly reject ungrammatical English sentences involving *wh*-movement out of relative clauses and adjunct clauses owing to L1 transfer. However, Grewendorf (1996, p.728) argues that if UG is accessible, L2 learners should be as successful as native speakers of English in rejecting *wh*-movement violations. He notes also that this is usually not the case.

White (1992) contends that even if scrambling of *wh*-phrases occurs in Japanese, the scrambling itself is subjected to a subjacency constraint[4], and therefore cannot be used explain L2 subjacency violations. Although not much research addresses *that*-trace effect influences from Japanese, research on Korean suggests that cross-linguistic influences should be similar.

In sum, if these claims and observations are correct, knowledge of Japanese may help Japanese learners of English to accurately reject English sentences involving subjacency and ECP violations. Our goal is to evaluate learners' L1 knowledge effects in light of the differential exposure to English before and after the critical period before modeling the influences of language learning aptitude. We take up this stage-wise comparison in the remainder of the paper.

To the extent that L1 transfer affects the learners' performance, it should apply consistently to all of the Japanese participants in our study. Variability in performance differences among the participants should be attributable to

three competing, and potentially interacting, sources: 1) L1 influences; 2) exposure to English during and after the critical period; 3) individual differences in aptitude for language analysis.

Method

Participants

As a preliminary step in designing the present study, a pilot study was conducted in the first of a four-year data gathering project. The pilot phase involved 75 Japanese university students who took all of the aptitude tests and a pilot grammaticality judgment task. From observed correlations between aptitude, group membership, and subjacency violation grammaticality judgements, effect sizes likely to be observed in a larger study were generated. The pilot study coefficients of determination (R^2s) were used as input into a statistical power simulation (Cohen, 1988; Cohen & Cohen, 1983). A power analysis provides parameter estimates for adequate sampling given a set of projected effect sizes. Since a non-rejection of the null hypothesis for the effect of aptitude on UG principles will lead to a substantive theory-based interpretation, it is essential that any insignificant effect size be interpreted as correctly indicating a true null hypothesis for the population. The power analysis simulation (Cohen & Bornstien, 1990) in this phase of the study indicated a minimum sample of 115 subjects in order to maintain a statistical power of .80. This became the sampling goal in the second year of the study.

One hundred twenty-nine Japanese university students participated in the main study. All students were between the ages of 18 and 21. As a proficiency baseline for comparing the groups, the Test of English as a Foreign Language (TOEFL) was used. In addition, an English language learning/exposure survey was conducted in order to provide data about individual differences in exposure to English. Membership in one of the three groups (levels of the independent variable) was determined by age of initial exposure to English, and length of residence overseas.

A control group of 17 native speakers of English was also formed. This group provided a baseline comparison for grammaticality judgements used as the criterion variables in the study.

Dependent variables

The criterion test in the study, the Wh-Movement Violation Survey, (WHVS) was devised to present subjects sentences manifesting three types of *wh*-movement in English. Different types of extractions were included in order to create violations and well-formed sentences. The subjects' task was to rate the degree of naturalness of each sentence.
Example:

What did Kate like the story that described?

-3	-2	-1	0	1	2	3
not natural						perfectly natural

The Wh-Movement Violation Survey consisted of 24 items. Twelve of the items constituted violations. Four were *that*-trace effect violations, four were movement out of relative clauses, and four were movement out of adjunct islands. The other 12 control items were well-formed examples of *wh*-movement in English. The survey was given to four groups; participants who had acquired English pre-critical period during residence overseas, a group who had acquired English through overseas immersion post-critical period, a post-critical period foreign language learner group who had never left Japan, and a group of native English speaker controls.

Operationalizing the Critical Period

Several putative critical periods have been used in SLA research, ranging from 7 through 16 years of age. In the present study, the cut-off age for operationalizing 'Child SLA' was set at the threshold age of puberty (less than age 12). As these participants had begun acquiring English well prior to the onset of maturation, we assume that they were the most influenced by UG principles. If a participant had lived in childhood in a country where English was the dominant language, or had attended an English-medium elementary school longer than three years before the age of twelve, he/she was classified as a child second language acquirer. Members of the Child SLA group had in general lived with their parents, who were posted on extensive overseas assignments with Japanese corporations.

Participants who were first exposed to English in adolescence, 'Teen SLA', we expect to have been comparatively less influenced by UG. The Teen SLA

classification was made if a participant had resided in an English-dominant country after the age of 12 for at least two years before returning to Japan at the age of 18 to attend college. This group had a shorter period of naturalistic exposure to English than the Child SLA group members.

For a third group who began learning English formally in middle school we anticipated the least relative influence by UG principles. The 'Teen FLL' classification was made if participants had never resided in an English-dominant country after the age of twelve or if they had never attended an English-medium international school in Japan. Members of this group had exposure to six years of English instruction beginning in junior high school. They were also enrolled in a teacher-certification program, and thus might be considered more highly motivated than average undergraduates. Few members of this group reported exposure to English outside of Japan, usually exposure through short-term summer vacations.

Table 1. Exposure group means

Group	N	TOEFL (SD)	Length of Residence
Child SLA/1	34	583 (36)	8.1
Teen SLA/2	38	548 (32)	3.2
Teen FLL/3	57	511 (43)	0.4

Table 1 provides the mean profile of English proficiency (TOEFL) and length of residence (LOR) overseas in years for each of the three groups participating in the study. At the time of the study all of the participants were between the ages of 18 and 21 years old.

The Child SLA group stands out in its pattern of exposure to English, with an average of 8 years of overseas residence. The mean TOEFL for this group is perhaps lower than what might be expected from an average of eight years of exposure. Since TOEFL content samples academic English, and many of the Child SLA group members had returned to Japan at junior or senior high school age, their TOEFL scores belie the near native-like fluency and accent that many of these students demonstrate in face to face interviews. The Teen SLA group do not have the same native-like fluency of the Child SLA group, mainly because they have developed their English proficiency in the context of immersion in junior or senior high school contexts. The Teen FLL group achieved a TOEFL score that is above the current national average (t =2.69, p <.01), which includes both overseas and non-immersion exposure to English Japanese learners (ETS, 1997), a fact which may indicate keenness and strong

motivation to get certification as future teachers, as well as exceptional talent in learning English as a foreign language.

Aptitude measurements

In addition to Test of English as a Foreign Language, the Japanese participants in this study took two aptitude tests in multiple choice format, which facilitated dichotomous (right/wrong) scoring and item analyses. The tests used were the MLAT — Modern Language Aptitude Test (Carroll & Sapon, 1959) Part 4–the Words in Sentences subtest, which was given in English. A Japanese-medium aptitude test was also given, the LABJ — Language Aptitude Battery for Japanese Part 2–the Artificial Language Analysis subtest (Sasaki, 1991, 1996). The Words in Sentences subtest of the MLAT was used as an index of learners' sensitivity to the semantic and syntactic roles that different constituents play in sentences. Example:

The player hit *the ball* with a bat.
The man left his hat on the bench
a b c d e

Here the task is to locate the semantic role in the test sentence that matches *the ball* in the cue sentence.

The rationale for using English as the medium of testing on one of the aptitude tests was to provide a possible counterbalance to the Japanese-medium aptitude measure used in the study. This decision was based on the possibility that some of the undergraduates who had acquired English as children could in fact be more literate in English than they would be in Japanese.

The LABJ was designed for adult Japanese language learners and has produced reliable data in Sasaki (1991, 1996). LABJ3 (artificial language) was designed to measure inductive language learning ability. It was based on a translation of PLAB-Part 4, where the examinees were given a gloss and two sentence structures of an unknown language as follows:

gade	father, a father
shi	horse, a horse
gade shir le	Father sees a horse
gade shir la	Father saw a horse
be	carries

After having studied the gloss and samples, the participants in the next 15

items had to choose a correct sentence in the artificial language corresponding to a target English sentence (e.g. "Father carries a horse"). For LABJ3, all glosses and sentences of the unknown language were changed into *katakana* (see Sasaki, 1991 for details). Other than the orthographic change, the method was the same as that used in PLAB-Part 4.

The use of two language analysis aptitude measures allows for a validity check relative to the participants' English language proficiency. Since English and Japanese-medium measures of aptitude are used, the effect of language-of-measurement can be assessed through correlational evidence. If the language-of-measurement creates an artefact, we would expect to find that language analysis aptitude measured via English would be more correlated with English proficiency than it would be with a Japanese-medium measure of aptitude. If the construct of aptitude is here language-of-measurement independent, we would expect that the two aptitude measures correlate with each other more than they would correlate with English proficiency. This issue is examined with the use of principal axis factor analysis below.

Reliability and validity evidence

The three tests used in the study were subjected to internal consistency reliability estimates (Kuder-Richardson 20 or Cronbach's alpha). All tests were Rasch-analyzed for fit and transformed onto a logit scale so as to provide a common metric. Table 2 provides a sketch of the internal consistency estimates for the tests used in the study.

Table 2. Internal consistency reliability estimates

Test	Code	Language	Reliability	Items
Words in sent.	MLAT4	English	.84	45
Artificial gram.	LABJ3	Japanese	.84	15
Wh-Movement Violation survey	WHVS	English	.83	24

In order to avoid a confound between ability and aptitude, the construct of aptitude as it is operationalized in this study must be distinct from the construct of proficiency. Given the fact that the data here were collected cross-sectionally *ex-post facto*, there is a possibility that aptitude could correlate with proficiency in an ambiguous manner. That is, proficiency could be the cumulative *result* of individual differences in earlier as well as current aptitude

states. Conversely, the construct of aptitude could be the consequence of individual differences in second language acquisition experience. Zero-order correlations between proficiency and aptitude (cf. Harley & Hart, 1997; DeKeyser, 2000) are often not straight forwardly interpretable for this reason. In order to avoid this type of potential conundrum, an exploratory factor analysis was conducted on the two aptitude measures and the three subtests which make up the TOEFL battery.

The factor analysis takes the matrix of correlations and searches for clusters of measures that are relatively interrelated. These are extracted and their communalities are estimated. Since the battery comprises five measures, we anticipate at least two factors if the three TOEFL proficiency measures indicate 'proficiency' and the two measures of aptitude covary to cluster into an 'aptitude' dimension.

Principal Axis Factor Analysis was used with an extraction criterion set at greater than 10% of the variance. Two factors were thus extracted as they accounted for 44 and 19 percent of the variance, respectively. Since all of the measured variables in some way assess language knowledge and would lead to correlated latent factors, an oblique rotation method was selected. Table 3 lists the factor loadings for the five measures.

Table 3. Factor loadings

Measure	PROFICIENCY	APTITUDE
TOEFL LC	0.737	-0.402
TOEFL ST	0.917	0.053
TOEFL RC	0.836	-0.071
LABJ3	-0.090	0.449
MLAT4	0.001	0.833

The exploratory factor analysis of the proficiency and aptitude measures suggests that they are distinct, but correlated latent constructs. The three measures of English language proficiency — TOEFL listening comprehension (LC), TOEFL grammar and vocabulary (ST), and TOEFL reading comprehension (RC) — all load highly on the proficiency factor. The two measures of aptitude, LABJ3 and MLAT4, show very small loadings on the proficiency factor. Although their loadings on the aptitude factor are not symmetric, they still suggest that aptitude and proficiency are distinct constructs.

Analyses of main effects

If complete access to UG principles governing *wh*-movement diminishes after the onset of puberty, there should be a distinct disadvantage for acquirers with exposure to English beginning in their teens. Two types of exposure are included in this study; exposure in an immersion environment (Child and Teen SLA), and exposure in a foreign language environment only (Teen FLL). We begin with global analyses of differences between the three groups on *that*-trace effect and subjacency violations compared with identification of well-formed *wh*-questions in English. We will then progress to differences in the type of violations before examining the effects of aptitude and its interaction with exposure.

The first test of the effects of critical period begins with a global classification of the 24 WHVS items into the 12 items that constitute violations of *that*-trace effect or subjacency with the other 12 that were well-formed examples of *wh*-movement in English. To this end, the ratings of grammaticality given by all of the Japanese participants were subjected to multivariate analysis of variance. Here, the two dependent variables are the sum of all well-formed items (ALLOK) and the sum of the twelve items that involve a violation of *that*-trace effect or subjacency (ALLVIO).

Table 4. MANOVA for effects of Critical Period on grammaticality judgements

UNIVARIATE F TESTS					
VARIABLE	SS	DF	MS	F	P
ALLOK	4.006	2	2.003	1.948	0.147
ERROR	116.169	113	1.028		
ALLVIO	1.440	2	0.720	0.876	0.419
ERROR	92.873	113	0.822		
MULTIVARIATE TEST STATISTICS					
WILKS' LAMBDA =	0.917				
F-STATISTIC =	2.471		DF = 4, 224	PROB = 0.045	

The main effect for the critical period (Table 4) is not large enough to suggest that the foreign language learners differ from the Teen SLA acquirers or the Child SLA acquirers in accepting the well-formed *wh*-questions. The test of violations likewise suggests that no group differences exist. This observation might at first glance be taken to support the 'full access' position (Epstein, Flynn & Martohardjono, 1996). At this stage such an interpretation is not war-

ranted, since we have not yet compared the Child SLA intuitions to a native speaker benchmark on the same items. Further, since there are different types of violations embedded in the 12 item index of *wh*-movement violation examples, there remains the possibility that group differences are hidden.

In order to examine this possibility, a second between-groups analysis is performed. In this phase, the means of the summed scores for all well-formed items and the means of the summed items containing a violation of the *that*-trace effect or subjacency principles are tested after entering aptitude covariates into the model. Thereafter, the means of each of the three *wh*-movement violation types are tested. Table 5 shows the main effects analysis for well-formed items.

Table 5. Multiple ANCOVA for effects of aptitude on well-formed items

ANALYSIS OF COVARIANCE					
SOURCE	SS	DF	MS	F	P
GROUP	5.092	2	2.546	2.506	0.087
MLAT4	0.142	1	0.142	0.139	0.710
LABJ3	0.000	1	0.000	0.000	0.992
APT	0.151	1	0.151	0.149	0.700
ERROR	105.664	104	1.016		

The influence of aptitude on the judgement of well-formed sentences appears to be negligible. There remains a slight trend (p=.087) favoring the Child SLA group in their mean recognition of the lack of any type of violation in the twelve well-formed sentences on the WHVS (Figure 1). However, these Japanese learners of English as a whole falter in their recognition of well-formed sentences (cf. White & Genesee, 1996). The tendency is to opt for the mid point of the Likert scale in the judgement process.

The situation changes when we consider the influence of the aptitude covariates on the summed ratings of *that*-trace effect and subjacency violations. Here, as the bottom portion of Figure 1 indicates, all of the Japanese groups tend in varying degrees to reject sentences containing violations. They do so, however, with less accuracy than does the native speaker control group.

When the aptitude covariates are entered into the model (Table 6), there is a clear effect of LABJ3 and the factor score APT, suggesting that rejections of sentences containing violations may be influenced by individual differences in aptitude for language analysis. The trend that indicated an advantage for the Child SLA group disappears entirely (p=.617).

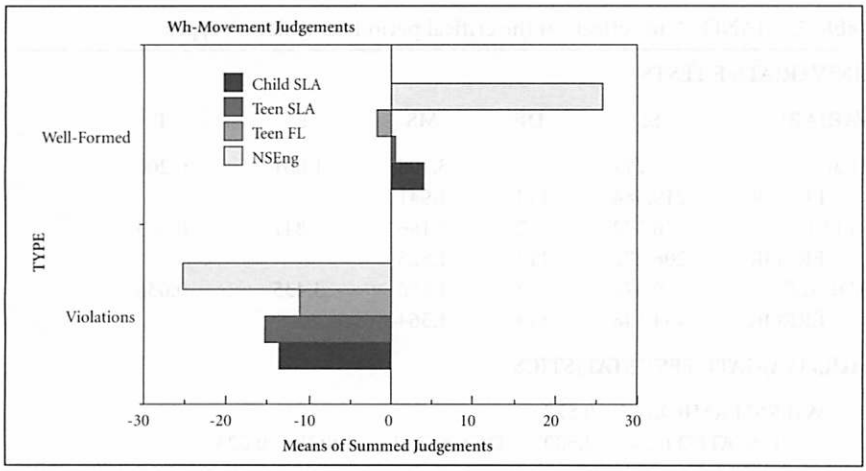

Figure 1. *Wh*-movement judgements for each group

Table 6. Multiple ANCOVA for effects of aptitude on violations

ANALYSIS OF COVARIANCE					
SOURCE	SS	DF	MS	F	P
GROUP	0.746	2	0.373	0.485	0.617
MLAT4	1.897	1	1.897	2.466	0.119
LABJ3	5.907	1	5.907	7.679	0.007
APT	6.430	1	6.430	8.359	0.005
ERROR	79.998	104	0.769		

Thus far we have observed that there is an effect for aptitude for language analysis that neutralizes the effect of exposure to English in childhood. The criterion variables have been the sum of well-formed item types and the sum of items containing violations of the *that*-trace effect or subjacency. In order to sort out specific effects, we now turn to testing the effects of aptitude and exposure on individual types of violations. We begin with a preliminary analysis to test the effect of exposure on violation types independently of aptitude. Here we of course omit the native speaker control group. Table 7 shows the effect of exposure grouping on each type of violation.

The groups do not appear to differ on their ratings of *that*-trace effect violations, but begin to separate progressively more on relative clause extractions and extractions out of adjunct islands. Raw score means (Figure 2) suggest that there are differential rejection rates for the three types of violations, with

Table 7. MANOVA for effects of the critical period on violation types

UNIVARIATE F TESTS					
VARIABLE	SS	DF	MS	F	P
That-trace	6.217	2	3.108	1.601	0.206
ERROR	219.384	113	1.941		
RELCL	10.372	2	5.186	2.841	0.063
ERROR	206.272	113	1.825		
ADJISLE	9.372	2	4.686	3.435	0.036
ERROR	154.148	113	1.364		

MULTIVARIATE TEST STATISTICS

WILKS' LAMBDA = 0.877
F-STATISTIC = 2.500 DF = 6, 222 PROB = 0.023

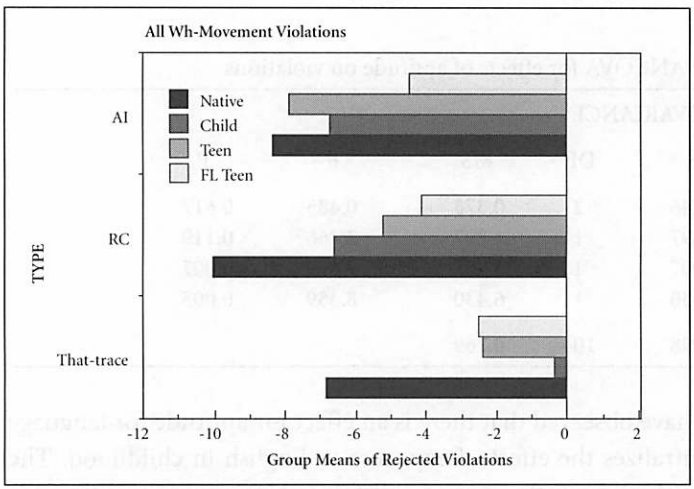

Figure 2. *Wh*-movement violations for each group

the *that*-trace effect being the least recognizable as a violation. We might infer at this point that there are only weak effects for post-critical period differences on the violation detection rates examined. We note that while the three groups differ among each other, they tend to differ in their judgements of grammaticality from the native speaker group on judgements of the *that*-trace effect violations and relative clause extraction, but less so on judgements of extractions from adjunct islands. At this stage, since the effects of aptitude have not yet been modeled, there remains the possibility that groups may differ in more

complex ways if individual difference factors work differentially *within* the three exposure profile groups.

Aptitude-exposure interactions

A notable phenomenon among proficiency, aptitude, and *wh*-movement tests in this study is the difference in between-group means, which are related to differences in exposure to English in childhood. The aptitude score mean differences in fact rival the mean differences seen in the ratings of *wh*-movement violations, suggesting that there may well be some other set of factors covarying with subjects' intuitions about grammaticality. As noted earlier, recent research has suggested that there may be other influences on post-critical period acquisition (Ioup, et al., 1994; White & Genesee, 1996). The question of interest here is whether individual differences in aptitude interact with the *that*-trace effect or subjacency domains.

Prior to assessing the influence of language learning aptitude measures on *wh*-movement phenomena, it is necessary to establish the extent of between-group differences on the two measures of aptitude. Here again, a multivariate analysis of variance is used.

Table 8. MANOVA effects for aptitude measures

UNIVARIATE F TESTS					
VARIABLE	SS	DF	MS	F	P
Artificial Grammar/LABJ3	7.662	2	3.831	3.921	0.023
ERROR	110.410	113	0.977		
Words in Sentences/MLAT4	26.013	2	13.006	19.657	0.000
ERROR	74.766	113	0.662		
MULTIVARIATE TEST STATISTICS					
WILKS' LAMBDA = 0.732					
F-STATISTIC = 9.449		DF = 4, 224		PROB = 0.000	

Significant main effects in Table 8 can be observed for both Artificial Grammar/LABJ3 and Words in Sentences/MLAT4. Of particular importance is the fact that both of these measures were designed to assess aptitude for the construct of language analysis. Figure 3 plots the group means for the *wh*-movement measures and the aptitude measures. The measures here are factor scores derived from a factor analysis of the three *wh*-movement measures and the two aptitude measures. The resulting loadings were rotated orthogonally

providing factor scores, which have been standardized (*z* scores) so as to show relative group mean outcomes on the grammaticality judgments and aptitude scores on a common scale.

The group factor score means strongly suggest that, in contrast with the age-related effects for *wh*-movement assessment, the advantage here goes to the foreign language learners. Two interpretations to account for why the foreign language learners obtained such relatively high aptitude scores come to mind: 1) the non-random sample of students comes from a teacher certification course which may tend to attract relatively 'talented' foreign language learners; 2) there is a 'transfer of training' from extensive form-focused instruction given in the current Japanese high school curriculum. Many students become familiar with the meta-language of English grammar before actually acquiring much tangible language proficiency. It is of interest to note that the majority of the Child SLA and Teen SLA groups did not graduate from high school in Japan, and in general tend to be relatively unfamiliar with metalinguistic terminology or methods of grammar analysis. The second of these interpretations is at variance with early research on aptitude which has asserted that language learning aptitude is not subject to amelioration through training or experience (see Skehan, 1998a for a review, and Sternberg, this volume for arguments and evidence in support of the view that aptitude *is* trainable and learnable).

The observed difference between Child SLA naturalistic acquirers and foreign language learners is striking. It may be taken as complementary evidence for what Harley and Hart (1997) and DeKeyser (2000) observed as near-zero correlations between their youngest second language acquirers relative to robust correlations between aptitude and proficiency among the older learners. In the present study, product-moment correlations between MLAT4, LABJ3 and TOEFL were -.151 and -.186 among the SLA Child group, and .411 and .195 among the Foreign Language Teens. The question we must turn to now is whether individual differences in aptitude influence the critical period-constrained grammaticality judgments of *wh*-movement phenomena.

In order to test the influence of the three aptitude measures (Words in Sentences/MLAT4, Artificial Grammar/LABJ3 and the factor scores APT) on subjects' relative ability to detect the three types of *wh*-movement violations, the general linear model (Tabachnick & Fidell, 1996) was used employing the two aptitude measures and the aptitude factor score as covariates and group codes as the independent variable. The three separate dependent variables in the analysis were the *that*-trace effect, relative clause extraction, and adjunct

island extraction subscores from the grammaticality judgment (WHVS) test. Here, the initial null hypothesis of interest is that the covariates (aptitude) will have no significant influence on the grammaticality judgments independently of the group codes, which, we hypothesize, encode the critical period effect. The multivariate main effect observed in Table 9 suggests that exposure profile effects remain even after we partial out the influence of the aptitude measures. The three types of *wh*-movement violations show varying effects related to the critical period, however. In order to examine the influence of aptitude

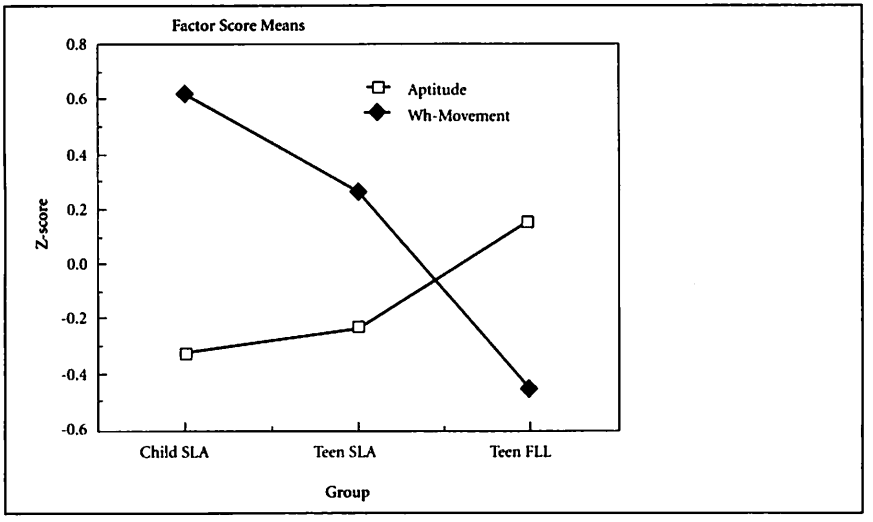

Figure 3. Aptitude and *Wh*-movement factor scores

Table 9. MANCOVA effects for exposure controlling for aptitude

UNIVARIATE F TESTS

VARIABLE	SS	DF	MS	F	P
That-Trace	4.608	2	2.304	1.167	0.315
ERROR	219.223	111	1.975		
RELCL	14.635	2	7.317	4.061	0.020
ERROR	199.984	111	1.802		
ADJISLE	12.810	2	6.405	4.720	0.011
ERROR	150.622	111	1.357		

MULTIVARIATE TEST STATISTICS

WILKS' LAMBDA = 0.856
F-STATISTIC = 2.927 DF = 6, 218 PROB = 0.009

on *that*-trace and subjacency violation detection independently, we need to test for interactions between aptitude and exposure to English. The interaction analysis permits a diagnosis of differential influence of aptitude on the grammaticality judgement process. It tests whether learners with relatively high language analysis aptitude (primarily FLL group members) do better in rejecting *wh*-movement violations independently of their status as members of the post-critical period exposure group.

The first step in this process is to examine each type of *wh*-movement violation separately in an ANCOVA model. The aim here is to test the assumption of the homogeneity of the regression planes (Tabachnick & Fidell, 1996), which will diagnose an interaction between critical period (exposure) effects and aptitude. In the event of a significant effect for the aptitude by group interaction, we continue with a Johnson-Neyman (JN) analysis of the interaction. The JN analysis (Huitema, 1980) allows us to locate the point on the covariate scale at which group differences become insignificant. That is, it diagnoses the amount of aptitude that allows post-critical period Teen FLL learners to detect *wh*-movement violations at a rate of accuracy comparable to the Child SLA subjects or the Teen SLA subjects.

Table 10. ANCOVA effects of aptitude on *that*-trace effect violations

ANALYSIS OF COVARIANCE/Test of homogeneity of regression slopes					
SOURCE	SS	DF	MS	F	P
GROUP	19.950	2	9.975	5.511	0.005
MLAT4	0.859	1	0.859	0.475	0.492
GROUP*MLAT4	20.291	2	10.146	5.606	0.005
ERROR	199.091	110	1.810		
ADJUSTED LEAST SQUARES MEANS					
	ADJ. LS MEAN		SE	N	
Child SLA	-0.140		0.323	29	
Teen SLA-	0.561		0.247	31	
Teen FLL-	0.876		0.208	56	

That-trace effect analysis

The first analysis is of the *that*-trace effect violations. It is noted that the three groups did not differ on their judgement of the *that*-trace effect violations at the outset. The present analysis tests the possibility that group differences are

obscured by differences in aptitude. If this hypothesis is correct, there will be a significant interaction between group membership and aptitude. As can be seen in Table 10, MLAT4 interacts with exposure, and in doing so reveals group differences in accuracy that were hidden earlier.

The scores adjusted for differences on the covariate (MLAT4) indicate that the Teen FLL group, which has the largest mean aptitude as measured by MLAT Words in Sentences, is most likely to reject the *that*-trace effect violation. The Child SLA group in general was the least keen to reject *that*-trace effect violations.

It appears that the *that*-trace effect is a 'weak' violation, which tends to be variable among native speakers of English (cf. Rizzi, 1990; White & Genesee, 1996), as indicated in Figure 2. While the Teen FLL participants with high aptitude rejected the violations of *that*-trace effect with the greatest relative accuracy, they are still dramatically different from the rejection rates of native speakers of English.

The Johnson-Neyman analysis indicates that Teen FLL participants with a logit of aptitude reaching 1.06 or larger will not differ significantly ($p < .05$) from the Child SLA subjects. It appears then, that once aptitude interacts with exposure to English, group differences do not reach significance using Bonferroni post-hoc criteria. Given the group means on aptitude (Appendix A), we estimate that about 60% of the Teen FLL learners reach the 'no difference' zone compared with the Child SLA group. We can infer then that aptitude allows these post-critical period foreign language teen acquirers to not differ from the average judgements of pre-critical period subjects. Still, when contrasted to the native speaker reference group, it appears that the three Japanese groups are equally unsure about the ungrammaticality of *that*-trace effect violations — all groups differ from the native speaker reference group on post-hoc tests. It would appear that neither exposure nor aptitude influences the acquisition of the *that*-trace effect for these learners of English.

Extraction out of Relative Clauses

Three measures of aptitude for grammatical analysis were used as covariates. Two, MLAT4 and LABJ3, were direct measures. The third measure was the aptitude factor score derived for each individual. Of the three aptitude measures, the factor score (APT) showed the largest covariance with the judgements of the grammaticality of the *wh*-movements from the relative clause position. It is therefore tested here for its interaction with group membership.

The test of homogeneity of regression slopes for relative clause extractions reveals that although aptitude (APT) shows a trend toward influencing judgements of grammaticality, it does not significantly interact with group membership.

Table 11. ANCOVA effects of aptitude on relative clause extractions

ANALYSIS OF COVARIANCE					
SOURCE	SS	DF	MS	F	P
GROUP	11.120	2	5.560	3.042	0.052
APT	5.394	1	5.394	2.951	0.089
APT*GROUP	1.921	2	0.960	0.525	0.593
ERROR	190.112	104	1.828		
ADJUSTED LEAST SQUARES MEANS					
	ADJ. LS MEAN	SE	N		
ChildSLA	-1.771	0.283	25		
TeenSLA	-1.384	0.263	29		
TeenFLL	-0.960	0.190	56		

Table 12. ANCOVA effects of aptitude on adjunct island extractions

ANALYSIS OF COVARIANCE					
SOURCE	SS	DF	MS	F	P
GROUP	14.441	2	7.220	5.088	0.007
LABJ3	0.831	1	0.831	0.585	0.445
LABJ3*GROUP	7.095	2	3.547	2.499	0.086
ERROR	212.880	150	1.419		
ADJUSTED LEAST SQUARES MEANS					
	ADJ. LS MEAN	SE	N		
ChildSLA	-1.653	0.217	32		
TeenSLA	-1.948	0.202	37		
TeenFLL	-1.047	0.133	56		

The group means, once slightly adjusted for differences in aptitude, show that the child SLA learners are most inclined to reject violations of subjacency when the movement is out of a relative clause. This finding generally conforms to the predictions made by the Critical Period Hypothesis among the

exposure profiles. Figure 2 indicates that the mean rejection rate for the Child SLA group approximates the mean observed for the native speaker reference group (p =.099). It appears here that aptitude has no influence on learners' detection of subjacency violations when the movement is out of relative clauses. Further, the argument that Japanese learners in general transfer a movement constraint from their L1 is not supported. However, there remains a significant difference between the native reference group and the post-critical period Teen SLA and Teen FLL groups (see Figure 2).

Adjunct island extractions

The analysis of differences in the judgements of the grammaticality of movement out of adjunct islands reveals a pattern different from the previous analyses. Here, neither the SLA Child group nor the Teen SLA group differ significantly from the native speaker reference group in rejecting subjacency violations when movement is out of adjunct islands. The question of interest here is whether the higher aptitude of the Teen FLL group interacts in a manner that decreases the between group differences.

The analysis of the interaction between exposure to English and aptitude indicates that there is a nearly significant effect. There is no significant main effect for aptitude directly on rejection of adjunct island extraction.

The Johnson-Neyman analysis conducted on the near-significant interaction (p=.086) indicates that Teen FLL group members with a LABJ3 logit score of >2.57 would not differ significantly from the Child SLA group with respect to the mean adjunct island rejection rate. Fifty-three percent of the foreign language learners meet this criterion with logit-scaled LABJ3 scores larger than 2.57, suggesting there is considerable influence of aptitude serving to compensate for a lack of exposure to English in childhood for these teen foreign language learners in recognizing subjacency condition violations involving adjunct island extractions.

Summary of results

Table 13 summarizes the main effects of the critical period, aptitude, and their interaction on the three types of *wh*-movement violations examined in this study. In the table, brackets [ ] denote ranges of significant post hoc differences among the exposure groupings relative to the native English speaker ref-

erence group, (NES). Groups outside of the brackets are significantly different ($p < .10$) from the native speaker reference group, while groups inside do not differ from each other once the influence of aptitude is modeled. Parentheses ( ) denote ranges of 'no difference' based on the Johnson-Neyman analyses. The percentage figure next to TFLL, the Teen foreign language group, indicates the percentage of that group with the amount of compensatory aptitude that, once modeled, results in no significant difference from the Child SLA (pre-critical period) group.

While there is an interaction between exposure to English before and after the end of the critical period and aptitude on the *that*-trace effect violations, even with the compensatory effect of aptitude, there is still a large difference between the native English reference group and all of these Japanese learners of English. Here, aptitude interacts with exposure, the result of which qualifies 60% of the foreign language learners as 'no different' from the Child SLA group in rejection rates of *that*-trace effect violations. Yet none of the Japanese groups approximates the rejection rate of the native English speakers.

Table 13. Summary of main effects ($p < .10$)and interactions

Test	Critical period		Aptitude	Interaction
That-Trace Effect	Yes NES	[(CSLA	No TSLA	Yes TFLL60%)]
Relative Clause	Yes [NES	CSLA]	Yes TSLA	No TFLL
Adjunct Island	Yes [NES	(CSLA	No TSLA]	Yes TFLL53%)

On the extraction from relative clause violations, there were main effects for the critical period influences and for aptitude. Here, the Child SLA group was nearest to the native English reference group mean (p =.099). Although there was a weak main effect for aptitude, it showed no interaction with group membership, obviating a need for the Johnson-Neyman analysis.

The adjunct island violation main effects analysis revealed that both the Child SLA and Teen SLA groups differed from the Teen FLL group as a whole, but did not differ from the native English reference group. Though not significantly different, the means of the Teen SLA group are in fact closer to the NES group than are the Child SLA group means. This result in particular suggests that some violations of *wh*-movement are detectable even for post-crit-

ical period acquirers, provided there has been exposure through immersion. Here again, aptitude interacts with exposure. The Johnson-Neyman analysis indicates that about half of the Teen FLL learners, those with the highest aptitude scores, are no different from the Child SLA group in rejecting these extractions from adjunct islands. We can infer also that the main effect result (done prior to the Johnson-Neyman analysis) is attributable to the Teen FLL with the least aptitude.

Conclusions

The first generalization to be made about the results of these analyses is that the effect of exposure before and after the critical period appears more complex than originally thought. The between-group differences, based on the age-of-acquisition grouping factor alone, indicate that the three groups are not equal in their accuracy in detecting violations of subjacency when extractions are from relative clauses or adjunct islands. Since all subjects are Japanese, transfer can be ruled out as a source of this difference, since rejection rates apparently vary with length of exposure to English. Moreover, among these Japanese groups, it appears that an interaction of aptitude and exposure induces a convergence effect that allows slightly more than half of the Teen FLL group members to approximate the rejection rates of the Child SLA and Teen SLA groups on the *that*-trace effect violations and extraction from adjunct islands type of violation.

If UG had been accessible to the SLA teens and FLL teens through their knowledge of Japanese, no main effects for exposure or aptitude would be expected. Knowledge of Japanese would be a sufficient basis for rejecting *wh*-movement violations in English. This was not the case.

As an alternative to a partial access mediated-by-exposure interpretation, we have provided empirical evidence that individual differences in language learning aptitude — particularly aptitude for language analysis — possibly indicate a factor that compensates for some learners' relatively late and infrequent access to English. Our findings support the notion of differential parameter resetting for post-critical period learners with high levels of metalinguistic awareness or aptitude for language analysis.

The effects of aptitude for language analysis are our approximation to Bley-Vroman's hypothesis (1989) that adult (post-critical period) problem solving — composed of cognitive processes such as analogical reasoning, dis-

tributional analysis, and hypothesis testing, is an important influence on post-critical period SLA. Whether language analysis, as we have operationalized it, conforms to Bley-Vroman's notion of 'problem solving traits' is a matter for further empirical exploration. Future studies of language learning aptitude need to be conducted longitudinally in order to examine whether the early emergence of individual differences endures across different profiles of exposure to an L2 in second and foreign language contexts.

We can surmise that a plausible picture of post-critical period SLA is a composite of varying influences which are likely to conform to both biological (critical period and possibly aptitude) and experiential (naturalistic exposure and formal learning) factors. As maturation-sensitive UG attenuates with age, differences in aptitude gradually emerge and interact with exposure to create variation in post-critical period SLA. The end result is the emergence of individual differences in peripheral *and* core grammar knowledge. We represent this notion in Figure 4. Here, darker shades indicate stronger influences relative to lighter shades.

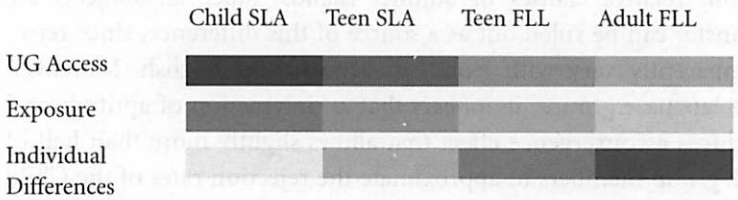

Figure 4. Maturational and experiential influences on SLA/FLL

The invariant influence on child second language acquisition, UG, (darkest cell) diminishes with maturation but interacts with exposure. Children in an SLA context have full access to UG and rich exposure (relatively dark cells) to positive evidence, but show the smallest variation (lightest shade) in individual differences (McLaughlin, 1985). Late critical period and early post-critical period teens show incipient maturation effects. Teens in an SLA context also have relatively robust exposure profiles, usually with access to both naturalistic exposure and formal language instruction. SLA teens, we surmise, will start to show the earliest emergence of individual differences in aptitude and metalinguistic awareness.

The situation for foreign language learners shows a dramatic difference. Teenage exposure to foreign language instruction is typically input-weak, based largely on deductive learning, and occurs near the end of the critical period. Individual differences here emerge to interact with the limited expo-

sure, which cumulatively leads to variance in acquisition. The weakest possible SLA profile in this scheme is that of the mature adult in a foreign language context, where diminished access to UG (other than that which is potentially transferable from L1) compounded with the weakest form of exposure (usually infrequent lessons), the end result is most dramatically different from target language norms. In the adult SLA profile, we assume that individual differences emerge as the greatest relative influence on variation in the acquisition of core and peripheral aspects of a second language grammar.

Finally, we have not addressed the possibility that what we have labeled 'aptitude' may be the *result* of extensive deductive learning experiences. Research on aptitude has largely operated under the assumption that it is immutable, though very little longitudinal research on language learning aptitude (cf. Skehan, 1986a, 1986b) has been undertaken. Skehan for instance infers that language learning aptitude, especially aptitude for verbal memory, may not be amenable to instruction. Our study cannot evaluate this interpretation, since our measures of aptitude were related to the construct of inductive language analysis. The results of this study do not necessarily support an interpretation of inductive language analysis ability being invariant over time. A plausible interpretation is that the teenage foreign language learners in this study appear to have relatively high aptitude because they have transferred extensive training in formal deductive grammar analysis to the grammaticality judgment tasks. The research agenda on the Critical Period in SLA needs to augment the growing body of work done on UG-based constraints with longitudinal studies of emergent aptitude after puberty in order to assess its role. SLA research needs, in short, to explore how 'exceptional' post-critical period second language learners ever manage to beat the odds.

Notes

* We would like to thank Kevin Gregg and an anonymous reviewer for comments on an earlier draft of the paper.

1. It should be noted that some scholars argue such constraints are not relevant to the *wh*-question formations such as *wh*-movements out of relative clauses (Deane, 1992).

2. In earlier work, the notion of bounding nodes is used. The subjacency condition holds that any movement crossing more than one bounding node (IP and NP in English) is blocked (Chomsky, 1977). Although Chomsky (1986, p.31) earlier posits that extraction out of an adjunct island involves crossing two barriers, in his later analysis (1986, p.66), this was reanalyzed when the moved element is an NP (e.g., *who*) as in:

a. he is the person who [IP they left [before speaking to *t*]] (before meeting *t*)

In this initial analysis, the matrix IP and the adjunct CP are the two barriers between *who* and its trace. However, in a later analysis, *who* adjoins to the *before*-phrase. This results in the movement of *who* out of the adjunct in (a) not crossing any barriers.

Lasnik and Saito (1992, p.91) say a sentence such as (b) "has the status of a weak Subjacency violation" .

b. [Which linguist]$_i$[IP did you [VP write your thesis] [after you consultedt_i]

In studies of second language acquisition, Martohardjono (1993, cited in Schwartz and Sprouse 2000, pp.176–177), White and Genesee (1996), and Shimizu (1994), treat these adjunct islands as strong violations.

3. Haig (2000) argues that subjacency does not constrain syntactic movements such as scrambling in Japanese. Rather, pragmatic constraints account for the unacceptable sentences that appear to violate subjacency.

4. Shimizu (1994) translated both English declaratives and interrogatives into Japanese in order to examine the status of Japanese sentences when *wh*-phrases are fronted, possibly by scrambling. He found that combined mean judgements of the translated sentences did not differ significantly from native English speaker judgements.

5. MANCOVA and ANCOVA impose a number of assumptions: normal distributions of covariates and the dependent variable; homogeneous regression slopes; reliability of covariates > .80; linearity as evidenced by predicted values of Y' plotted against residuals after adjusting for the effects of covariates. All of these assumptions were met in this data.

Appendix A

The following results are for:
Child SLA Group

Total observations: 34

	LABJ3	MLAT4	APT
N OF CASES	32	31	25
MEAN	2.306	0.286	-0.221
VARIANCE	1.646	0.792	1.029
STANDARD DEV	1.283	0.890	1.014

The following results are for:
Teen SLA Group

Total observations: 38

	LABJ3	MLAT4	APT
N OF CASES	37	32	29
MEAN	2.347	0.875	-0.213
VARIANCE	1.085	0.674	0.971
STANDARD DEV	1.041	0.821	0.985

The following results are for:
Teen FLL Group

Total observations: 57

	LABJ3	MLAT4	APT
N OF CASES	57	57	56
MEAN	2.838	1.504	0.379
VARIANCE	0.577	0.603	0.846
STANDARD DEV	0.760	0.777	0.920

In the series LANGUAGE LEARNING & LANGUAGE TEACHING (LL&LT) the following titles have been published thus far, or are scheduled for publication:

1. CHUN, Dorothy M.: *Discourse Intonation in L2. From theory and research to practice.* 2002.
2. ROBINSON, Peter (ed.): *Individual Differences and Instructed Language Learning.* 2002.
3. PORTE, Graeme Keith: *Appraising Research in Second Language Learning. A practical approach to critical analysis of quantitative research.* 2002.
4. TRAPPES-LOMAX, Hugh and Gibson FERGUSON: *Language in Language Teacher Education.* 2002.
5. GASS, Susan, Kathleen BARDOVI-HARLIG, Sally Sieloff MAGNAN and Joel WALZ (eds.): *Pedagogical Norms for Second and Foreign Language Learning and Teaching.* 2002.
6. GRANGER, Sylviane, Joseph HUNG and Stephanie PETCH-TYSON (eds.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching.* 2002.