The Impact of Early English Exposure/Education on Vocabulary Size Through Bayesian Hierarchical Modeling: An Additional Analysis of Katagiri (2019)

Kazuhiko Katagiri

1. Introduction

Dozens of studies have investigated the subsequent effects of Early Exposure to English and/or Early English Education (EEE) in preschool or elementary school. However, few studies have focused on the effects manifested in receptive vocabulary knowledge (receptive vocabulary size, the number of English words that learners know) in detail. In particular, few previous studies have investigated the effects that appear in receptive vocabulary knowledge over three years after becoming a high school student longitudinally (with repeated measurements), as far as the author has investigated, except for a previous study by the author, Katagiri (2019).

From April 2020, with the implementation of the new Course of Study (Learning Guideline) for Elementary School, which was revised and released in December 2017 by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), foreign language activities started from the third grade in all elementary schools in Japan. Learning English as an official subject also began in the 5th and 6th grades in all elementary schools. In fact, many elementary schools were allowed to implement these measures in advance from the 2018 academic year (MEXT, 2017). It has become impossible to conduct domestic research targeting

Japanese native speakers receiving education in Japan, where new high school students who have not received early English education inside or outside the elementary school are newly found and new data are collected and researched.

In Katagiri (2019), two main areas were investigated. Firstly, the proportion of infants and children in Japan in the 1990s who received early English education inside and outside kindergarten and elementary school was examined. Secondly, the subsequent effects of receiving early English education during pre-kindergarten, kindergarten, and/or elementary school days were focused on, specifically the effects that appear in receptive vocabulary knowledge (vocabulary size). The effects after adolescence were investigated over three years after becoming a high school student. The method of Katagiri (2019) involved the same high school students taking the receptive vocabulary size test four times in three years. The measurement results of each time point were divided into two groups, "high school students who did not receive early English education (Non-EX)" and "high school students who received it (EX)" (for data analysis). Bayesian t-tests were conducted at each time point, and the evaluation (judgment) was conducted with the numbers of "Bayes Factors" and "effect sizes (Cohen's *d*)." From the results of Katagiri (2019), the following six possibilities were found:

(1) Early English education was conducted in a variety of ways and at a considerable rate even in regional cities that are not metropolises in the 1990s. The percentage of early English education experience inside and/or outside kindergarten and nursery school was high at 41.9%. The percentage of students who had experience with English inside and/or outside school in the first and second grades of elementary school was 23.3%, while those who had experience in the third and fourth grades was 31.2%, and those who had experience in the fifth and sixth grades was 59.9%. The percentage of early English education experience during the elementary school days increased as the grades progressed from lower to higher.

(2) If the group of high school students had experience with English in the first and second grades of elementary school, they knew a few more English

words at the time of high school entrance than the other group that did not have experience. In other words, it was shown that the effect of having been exposed to English in the first and second grades of elementary school remained and appeared until the time of high school entrance.

(3) If the group had experience with English in the third and fourth grades of elementary school, they knew a few more English words at the time of high school entrance, the end of high school first year, the end of high school second year, and the end of high school third year than the other group that did not have experience. In other words, it was shown that the effect of having been exposed to English in the third and fourth grades of elementary school remained and appeared throughout the time from high school entrance to high school graduation.

(4) If the group had experience with English in the fifth and sixth grades of elementary school, they knew a few more English words at the time of entrance and the end of the first year of high school than the other group that did not have experience. In other words, it was shown that the effect of having been exposed to English in the fifth and sixth grades of elementary school remained and appeared until the end of the first year of high school.

(5) If the group had experience with English from kindergarten and nursery school to the fifth and sixth grades of elementary school, their progress of receptive vocabulary knowledge tended to be slightly larger than the progress of the other group that did not have experience (the effect size is "small"). However, the Bayesian factor was not sufficient to confirm this.

(6) From the results of high school students who had English classes at the elementary school they attended (students who spent the fourth to sixth grades of elementary school between 1998 and 2000) from the questionnaire description, the "effect of the dawn of elementary school English education (1998-2000)" was indicated.

On the other hand, for data with hierarchy, it is appropriate to estimate using a hierarchical model, and analysis by a hierarchical model has been recommended especially for approximately the last decade (Haebara, 2014; Hox, 2010; Komori, 2022; Kubo, 2012; Ozaki et al., 2018; Shimizu, 2014). Although it has not become popular in Second Language (L2) Teaching/Learning Research yet, it has started to be used. As shown in Figure 1 described later, the data of Katagiri (2019) is hierarchical data.

Therefore, although the purposes of this study and Katagiri (2019) are not exactly the same, their overall frameworks are very similar, and the data uses exactly the same data. However, the statistical analysis method is “hierarchical model analysis” in this study, and the purpose is to obtain new insights. When conducting “hierarchical model analysis,” both “Bayesian hierarchical model” using “Bayesian approach through Markov chain Monte Carlo (MCMC) estimation method” and “hierarchical model by maximum likelihood (ML) estimation method” using traditional statistics “ML estimation method” are conducted at the same time. The difference in analysis output results is also considered from the research perspective of the stability (reliability) of statistical estimation results. Katagiri (2019) was a study that conducted “a stack of micro-analyses” that compared Non-EX and EX at each time point where the same high school students took the receptive vocabulary size test four times in three years. In contrast, this study conducts “macro analysis only once” by conducting hierarchical model analysis (with EX being further divided into five groups according to the period when the students were exposed to English at their early ages).

From the questionnaire results of Katagiri (2019), it was found that the term “early exposure with English (experience of being exposed to English inside and outside kindergarten and/or school in early childhood and elementary school)” is more representative of reality than “early English education.” Therefore, the abbreviation of “EEE” indicates both “Early Exposure with English” and “Early English Education” in this paper.

2. Research Design

2.1 Purpose

The purpose of this study is: (1) to investigate whether the group(s) that self-reported having experienced EEE inside or outside the kindergarten and/or school during early childhood or elementary school for a longer time (being divided into 5 groups based on the length of time they experienced EEE) has a larger receptive vocabulary knowledge size (i.e., knows more English words) during the three years of high school learning period, compared with the group that self-reported not having experienced it through longitudinal measurement of receptive vocabulary knowledge over three years of high school, and (2) to investigate whether the receptive vocabulary knowledge grows more over three years of high school, and (3) to investigate the differences in statistical analysis output results of “Bayesian hierarchical model by way of MCMC estimation” and “hierarchical model by way of ML estimation.”

2.2 Participants

252 students (one grade, all 7 classes) who entered S Prefectural High School in April 2004 participated in the academic years 2004-2006. These are students who spent their elementary school years from 1995 to 2000. In other words, these were students who had studied at elementary school before 2002 when “*sougou tekina gakusyu no jikan* [comprehensive learning time]” was introduced in all elementary schools in Japan. This was a period when children had little opportunity to get familiar with English. They are the same participants as in Katagiri (2019), as well as in Katagiri (2009, 2010).

2.3 Materials

2.3.1 English Learning History Questionnaire

An “English Learning History Questionnaire” was created to investigate early English education experience. The questionnaire consists of 12 items. For Q1 to Q5, we asked about the experience of listening to English songs, playing English games, reading English picture books, using English materials, and having English conversations, respectively. The questions were asked for the following periods: "Q1 before entering kindergarten or nursery school, Q2 before entering elementary school, Q3 when in the first and second grades of elementary school, Q4 when in the third and fourth grades of elementary school, Q5 when in the fifth and sixth grades of elementary school." Free descriptive answers were also allowed for each question.

2.3.2 Receptive Vocabulary Size Measurement Test

As a test to measure the receptive vocabulary knowledge of Japanese native speakers learning English, the Mochizuki's (1998) Vocabulary Size Test was used. Three versions of the test are available, but all three versions (version 1, version 2, and version 3) were used. The validity and reliability of the test as a receptive vocabulary size measurement tool (Katagiri, 2007) and that these three versions are parallel tests to each other (Katagiri, 2002) have been verified.

2.4. Procedures

The participants completed the Vocabulary Size Test at various points throughout their senior high school years. They took version 1 in April of the first year (upon entering senior high school), version 2 in January of the first year (approximately at the end of the first year), version 3 in January of the second year (approximately at the end of the second year), and version 1 again in November of the third year (approximately at the end of the third year). The first time in April of the first year and the last time at the end of November of the third year were the same version (ver. 1), but the test question papers were collected, and there is a long time of about 2 years and 8 months between the test implementation dates of both. The test does not have elements that are easy to remember, such as "content" and "story (outline of the story)," so the "learning effect" and "repetition effect" due to taking the same version (ver. 1) are to be discarded. Note that the data obtained by these vocabulary size tests is the same as the data obtained in Katagiri (2009).

The author will convert each response to the questions (from Q1 to Q5) in the English Learning History Questionnaire on EEE into a value of 1 for "yes" and 0 for "no," and will create a variable called "Exposed/Educated Periods (abbreviated as "ExpPrd")," which is the subtotal for each student from Q1 to Q5. Consequently, this variable (ExpPrd) has six levels: 0, 1, 2, 3, 4, or 5. ExpPrd 1 indicates that the participants had less than or equal to two years of EEE, ExpPrd 2 indicates more than two but less than or equal to four years of EEE, ExpPrd 3 indicates more than four but less than or equal to six years of EEE, ExpPrd 4 indicates more than six but less than or equal to eight years of EEE, and ExpPrd 5 indicates more than eight but less than or equal to ten years of EEE during their pre-school or elementary school days. ExpPrd 0 indicates that they had no EEE at all.

2.5 Data Analysis

Various point estimates of receptive vocabulary sizes for each group at each time will be estimated and described. Then, to see the overall trend, the marginal averages will be analzed.

This study's data form a hierarchy with "corresponding longitudinal data" at Level 1 and "control group, experimental group 1, experimental group 2… experimental group 5 (explanatory variable = independent variable)" at Level 2, as shown in Figure 1. In this analysis, although our main analysis divides the "experienced" group into five groups according to their periods of experience at Level 2 in addition to the "non-experienced" group for statistical processing as described later, Figure 1 shows a simplified diagram representing "experienced (EX) vs. non-experienced (Non-EX)" for ease of understanding due to space constraints.

Figure 2 shows the mathematical model for "the hierarchical modeling with random intercepts and random slopes" (Ozaki et al., 2019; Komori, 2022). In the equations shown in Figure 2, subscript i represents measurement points (i = 0, 1, 2, 3), where i = 0 represents the first test and i = 3 represents the fourth test. The author set up the variable data for measurement points so that the average vocabulary size test score at the time of the first test becomes the "intercept," and by doing so, the meaning of "slope" becomes the increase rate until the next test. Subscript j represents student ID numbers (j = 1, 2, 3, …, 252). At Level 1, in Equation (1.1), Vocabulary_{ij} represents the vocabulary size test scores for the student (subscript j) at the measurement point (subscript i), and is the dependent variable for the entire mathematical model described in Figure 2. Time_{ij} represents the Time variable for the student (subscript j) at the measurement point (subscript i), and is the secondary independent variable in Figure 2. β_{0j} represents intercepts, β_{1j} represents slopes, and ε_{ij} represents deviation (residual error) for the entire mathematical model. The variable ε_{ij} in Equation (1.2) means that it follows a normal distribution with a mean of zero and a standard deviation of σ (Sigma). At Level 2, ExpPrd_j in Equations (1.3) and (1.4) represents ExpPrd for the student (subscript j). This is a primary independent variable for the entire mathematical model described in Figure 2. γ_{00} and γ_{10} in Equations (1.3) and (1.4) represent their respective intercepts; γ_{01} and γ_{11} represent their respective slopes; u_{0j} and u_{1j} represent their respective deviations (errors). The two-variable vertical vector $(u_{0j}, u_{1j})'$ in Equation (1.5) means that it follows a bivariate normal distribution with a mean vector of zero and a variance-covariance matrix $\boldsymbol{T}$ in Equation (1.6). In matrix $\boldsymbol{T}$ in Equation (1.6), $\tau_{00}{}^2$ represents the variance related to u_{0j}, $\tau_{11}{}^2$ represents the variance related to u_{1j}, and τ_{01} represents the covariance between τ_{00} and τ_{11}. Consequently, r in Equation (1.7)

Figure 1
The Simplified Hierarchical Modeling of This Research Data

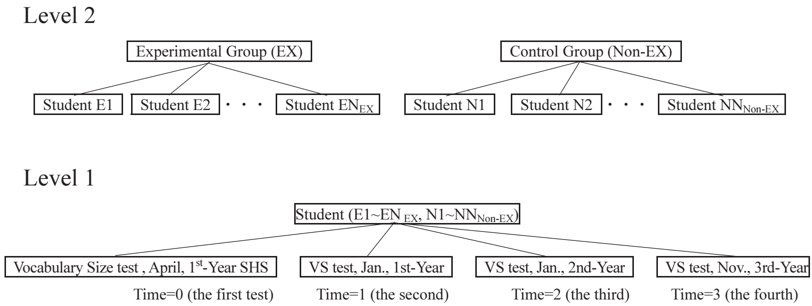


Figure 2

The Mathematical Model for the Hierarchical Modeling for This Research Data

Level 1 :

$$\text{Vocabulary}_{ij} = \beta_{0j} + \beta_{1j}\,\text{Time}_{ij} + \varepsilon_{ij} \tag{1.1}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \tag{1.2}$$

Level 2 :

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{ExpPrd}_j + u_{0j} \tag{1.3}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{ExpPrd}_j + u_{1j} \tag{1.4}$$

$$(u_{0j}, u_{0j})' \sim MVN(\mathbf{0}, \boldsymbol{T}) \tag{1.5}$$

$$\boldsymbol{T} = \begin{bmatrix} \tau_{00}{}^2 & \tau_{01} \\ \tau_{01} & \tau_{11}{}^2 \end{bmatrix} \tag{1.6}$$

$$r = \tau_{01} / (\tau_{00} * \tau_{11}) \tag{1.7}$$

indicates Pearson's product-moment correlation coefficient between τ_{00} and τ_{11}. The r in Equation (1.7) also indicates the partial correlation between β_{0j} (intercept) and β_{1j} (slope), which excludes the influence of ExpPrd$j$. Figure 2 indicates that both the intercept and slope in Equation (1.1) are random variables. "The hierarchical modeling with random intercepts and random slopes" shown in Figure 2 will be utilized in this study.

Statistical analyses will be conducted utilizing R (R Core Team, 2023) and RStudio (RStudio Team, 2023). Among probabilistic programming languages, RStan (Stan Development Team, 2023a) will be used in this study. RStan is the R interface to Stan. Stan uses two Markov chain Monte Carlo (MCMC) algorithms: the Hamiltonian Monte Carlo (HMC) algorithm and its adaptive variant—the no-U-turn sampler (NUTS) (Stan Development Team, 2023b). The hierarchical modeling shown in Figure 2 will be estimated through both Bayesian estimation and ML estimation utilizing both "stan_lmer function" in the rstanarm package (Stan Development Team 2023a) and "lmer function" in the lme4 package (Bates et al., 2015) for R, respectively. In Bayesian estimation, following Toyoda (2017), MCMC samplings will be generated with a set of "prior distributions = default settings in R (the details will be described in each estimation in the next chapter)", "chains = 5", "iterations = 21,000", "warmup (burn-in) = 1,000", and "thinning = none," in each estimation.

3. Results and Discussion

3.1 The Questionnaire

The results of the questionnaire analyses were reported in detail by Katagiri (2019), so they are not reprinted here. The frequencies of the ExpPrd are shown in Tables 1-1 and 2-1.

3.2 Numerical Summaries of Generated Quantities

Various estimates of generated quantities were obtained using MCMC (HMC algorithm and NUTS) with a wide range of a continuous uniform distribution as the prior distribution (a default setting in RStan [Stan Development Team, 2023a]). For parameters estimated by MCMC, this study uses an expected a posteriori (EAP) and a Sigma, which respectively represent a mean (μ) and a standard deviation (σ) of the posterior predictive distribution, as point estimates. A "post.sd" is the abbreviation for a posterior standard deviation, which is similar to the standard error (SE) in frequentist statistics. Thus, "EAP post.sd" is the posterior standard deviation in estimating the EAP from MCMC sampling. "Sigma post.sd" is the posterior standard deviation in estimating the Sigma from MCMC sampling. "EAP 2.5%" indicates the lowest level of the 95% Credible Interval (CrI) of the EAP estimate, and "EAP 97.5%" shows the highest level of the 95% CrI. "Sigma 2.5%" indicates the lowest level of the 95% CrI of the Sigma estimate, and "Sigma 97.5%" shows the highest level of the 95% CrI. Various point estimates of receptive vocabulary sizes for each group at each time through MCMC sampling are shown in Table 1-1. The indicators for examining the convergence of their estimates through MCMC sampling are shown in Table 1-2. The indicator "n_{eff} (n_eff)" shows the effective sample size, and "n_{eff}/N ratio" represents the ratio of the number of effective samples to all the samples. The indicator “mcse” shows the Monte Carlo Standard Error, and “mcse/post.sd” represents the ratio of the mcse to the posterior standard deviation. The indicator “rhat” shows the ratio of inter-chain variance to intra-chain variance. If “n_{eff}/N ratio” does not satisfy “0.1 or more”, if “mcse/post.sd” does not satisfy “0.1 or less”, or if “rhat” does not satisfy “1.1 or less”, the estimation through MCMC sampling may not be considered “converged” (Suzuki et al., 2020; Matsuishi et al., 2020). No problems related to convergence were found through the indicators of “n_{eff}/N ratios,” “mcse/post.sd,” and “rhats,” which are shown in Table 1-2. The line graph of receptive vocabulary sizes for each group at each time over three years of high school with 95% CrIs is shown in Figure 3. Trends can be

Table 1-1

Various Point Estimates of Receptive Vocabulary Sizes for Each Group at Each Time Through MCMC Sampling

Time	ExpPrd	N	**EAP (μ)**	EAP post.sd	EAP 2.5%	EAP 97.5%	**Sigma EAP**	Sigma post.sd	Sigma 2.5%	Sigma 97.5%
0 (first)	0	56	2,462	46	2,372	2,551	342	33.7	276	408
	1	81	2,339	41	2,259	2,419	365	29.3	308	422
	2	56	2,630	66	2,500	2,760	495	48.6	399	590
	3	26	2,586	78	2,433	2,739	394	59.8	277	512
	4	21	2,821	164	2,499	3,143	744	128	492	995
	5	12	2,600	137	2,332	2,868	462	116	234	690
1 (second)	0	56	3,113	78	2,961	3,265	578	56.8	467	690
	1	81	3,090	55	2,983	3,197	489	39.3	412	566
	2	56	3,331	86	3,163	3,500	642	63.1	518	766
	3	26	3,278	72	3,138	3,419	360	54.5	253	467
	4	21	3,525	174	3,184	3,865	782	135	518	1,046
	5	12	3,397	152	3,099	3,696	511	130	256	767
2 (third)	0	56	3,340	97	3,150	3,530	724	71.1	585	864
	1	81	3,111	68	2,979	3,243	608	49.1	511	704
	2	56	3,461	93	3,279	3,642	689	67.7	556	822
	3	26	3,494	127	3,245	3,743	638	97.2	448	829
	4	21	3,647	229	3,199	4,095	1,034	180	682	1,386
	5	12	3,725	166	3,399	4,051	560	142	282	837
3 (fourth)	0	56	3,649	127	3,401	3,897	949	93	767	1,131
	1	81	3,550	99	3,355	3,745	892	72.1	751	1,034
	2	56	3,861	130	3,605	4,116	973	95.7	786	1,161
	3	26	4,032	157	3,724	4,341	795	121	558	1,031
	4	21	4,170	281	3,620	4,721	1,265	218	837	1,693
	5	12	4,309	272	3,776	4,842	915	232	461	1,369

observed from Table 1-1 and Figure 3 that ExpPrd 1, who experienced EEE for less than 2 years, is slightly lower than Non-EX (ExpPrd 0), who had no experiences of EEE, and that ExpPrds 2, 3, 4, and 5, who experienced EEE for 2 years or more, are higher than Non-EX (ExpPrd 0). However, these tendencies will be examined in more detail later with the discussion of the results of the following statistical analyses.

Table 1-2

The Indicators for the Convergence of the Estimates for Table 1-1 Through MCMC Sampling

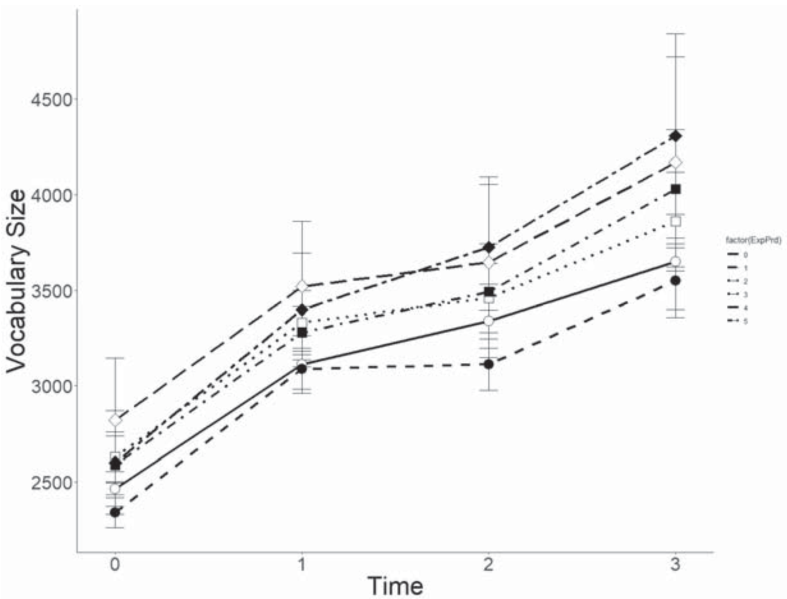
Time	ExpPrd	N	n_{eff}	n_{eff}	n_{eff}/N ratio	n_{eff}/N ratio	mcse	mcse	mcse/ post.sd	mcse/ post.sd	rhat	rhat
			as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma
0 (first)	0	56	73,925	75,857	0.739	0.759	0.172	0.126	0.004	0.004	1	1
	1	81	75,886	75,617	0.759	0.756	0.147	0.112	0.004	0.004	1	1
	2	56	80,568	71,896	0.806	0.719	0.241	0.193	0.004	0.004	1	1
	3	26	66,794	60,537	0.668	0.605	0.315	0.245	0.004	0.004	1	1
	4	21	65,124	59,628	0.651	0.596	0.622	0.523	0.004	0.004	1	1
	5	12	49,869	47,299	0.499	0.473	0.603	0.541	0.004	0.005	1	1
1 (second)	0	56	69,696	75,330	0.697	0.753	0.274	0.211	0.004	0.004	1	1
	1	81	76,743	73,760	0.767	0.738	0.185	0.148	0.003	0.004	1	1
	2	56	82,295	74,831	0.823	0.748	0.294	0.222	0.003	0.004	1	1
	3	26	67,536	70,272	0.675	0.703	0.256	0.207	0.004	0.004	1	1
	4	21	56,720	55,781	0.567	0.558	0.637	0.551	0.004	0.004	1	1
	5	12	48,068	45,123	0.481	0.451	0.676	0.595	0.004	0.005	1	1
2 (third)	0	56	69,132	75,113	0.691	0.751	0.357	0.271	0.004	0.004	1	1
	1	81	78,304	81,843	0.783	0.818	0.224	0.172	0.003	0.004	1	1
	2	56	81,020	84,451	0.810	0.845	0.367	0.248	0.004	0.004	1	1
	3	26	76,066	65,778	0.761	0.658	0.456	0.37	0.004	0.004	1	1
	4	21	71,629	63,132	0.716	0.631	0.925	0.731	0.004	0.004	1	1
	5	12	51,573	45,285	0.516	0.453	0.734	0.653	0.004	0.005	1	1
3 (fourth)	0	56	76,631	77,089	0.766	0.771	0.437	0.338	0.003	0.004	1	1
	1	81	85,548	78,224	0.855	0.782	0.338	0.248	0.003	0.003	1	1
	2	56	75,457	69,040	0.755	0.690	0.488	0.372	0.004	0.004	1	1
	3	26	64,076	58,311	0.641	0.583	0.568	0.486	0.004	0.004	1	1
	4	21	61,711	55,799	0.617	0.558	1.12	0.954	0.004	0.004	1	1
	5	12	52,769	45,096	0.528	0.451	1.22	1.05	0.004	0.005	1	1

3.3 The Marginal Means Estimated Through MCMC sampling

The marginal means for each group, which neglect the influence of the "Time" variable, were estimated through MCMC sampling with a wide range of a continuous uniform distribution as the prior distribution (a default setting in RStan [Stan Development Team, 2023a]) and are shown in Table 2-1; the indicators for examining the convergence of their estimates through MCMC sampling are shown in Table 2-2. The traceplots of MCMC samplings for the μ (mu) parameters of the six groups (ExpPrd = 0, 1, 2, 3, 4, 5) are shown in Figures 4-1, 4-2, 4-3, 4-4, 4-5, and 4-6, respectively. No problems related to the convergence in estimating the marginal means for each group through MCMC sampling were found from Table

Figure 3

The Line Graph of Vocabulary Sizes Over Three Years of High School with 95 % Credible Intervals



Note: ○=ExpPrd 0, ●=ExpPrd 1, □=ExpPrd 2, ■=ExpPrd 3, ◇=ExpPrd 4, ◆=ExpPrd 5

Table 2-1
Various Point Estimates of Vocabulary Sizes for Each Group at All Four Times Through MCMC Sampling (Marginal Means)

Time	ExpPrd	N	**EAP (μ)**	EAP post.sd	EAP 2.5%	EAP 97.5%	**Sigma EAP**	Sigma post.sd	Sigma 2.5%	Sigma 97.5%
ALL	0	56	3,141	53.3	3,037	3,246	800	38.3	725	875
	1	81	3,022	41.9	2,940	3,105	750	29.7	692	809
	2	56	3,321	56.0	3,211	3,430	834	39.6	756	911
	3	26	3,347	74.3	3,202	3,493	758	53.4	653	862
	4	21	3,539	114.0	3,316	3,762	1,036	82.2	875	1,198
	5	12	3,507	123.0	3,266	3,748	850	90.7	672	1,027

Table 2-2
The Indicators for the Convergence of the Estimates for Table 2-1 Through MCMC Sampling

Time	ExpPrd	N	n_{eff}	n_{eff}	n_{eff}/N ratio	n_{eff}/N ratio	mcse	mcse	mcse/ post.sd	mcse/ post.sd	rhat	rhat
			as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma	as to EAP	as to Sigma
ALL	0	56	95,309	95,130	0.953	0.951	0.182	0.126	0.003	0.003	1	1
	1	81	80,481	79,487	0.805	0.795	0.152	0.104	0.004	0.004	1	1
	2	56	78,400	80,143	0.784	0.801	0.196	0.13	0.004	0.003	1	1
	3	26	79,832	78,020	0.798	0.780	0.285	0.204	0.004	0.004	1	1
	4	21	82,778	70,516	0.828	0.705	0.404	0.311	0.004	0.004	1	1
	5	12	72,448	76,672	0.724	0.767	0.453	0.334	0.004	0.004	1	1

2-2 and Figures 4-1 through 4-6. The posterior distributions (the violinplot of the estimates of the marginal means through MCMC sampling) are displayed in Figure 5. Then, the probabilities of dominance among the marginal means and their effect sizes (Cohen's *d*, which is shown in Equation [2]) were estimated through MCMC sampling and are displayed in Table 3.

$$d = |(Mean_1 - Mean_2) / (\mathrm{sqrt}(((n_1 * (SD_1)^2) + (n_2 * (SD_2)^2))/(n_1 + n_2)))| \qquad (2)$$

Not Plonsky and Oswald's (2014) criterion but the classical Cohen's (1988) criterion is utilized (small effect size: $0.2 \leqq d < 0.5$, medium: $0.5 \leqq d < 0.8$, large: $0.8 \leqq d$) in this study.

ExpPrd 1 is lower than Non-EX (ExpPrd 0) with a probability of 96.0%, an EAP (μ) of Cohen's *d* of 0.15, a "post.sd" of Cohen's *d* of (0.09), and a 95% CrI of

Figures 4-1, 4-2, 4-3, 4-4, 4-5, and 4-6

The Traceplots of MCMC Sampling for the μ (Mu) Parameters of the Six Groups (ExpPrd = 0, 1, 2, 3, 4, 5), Respectively

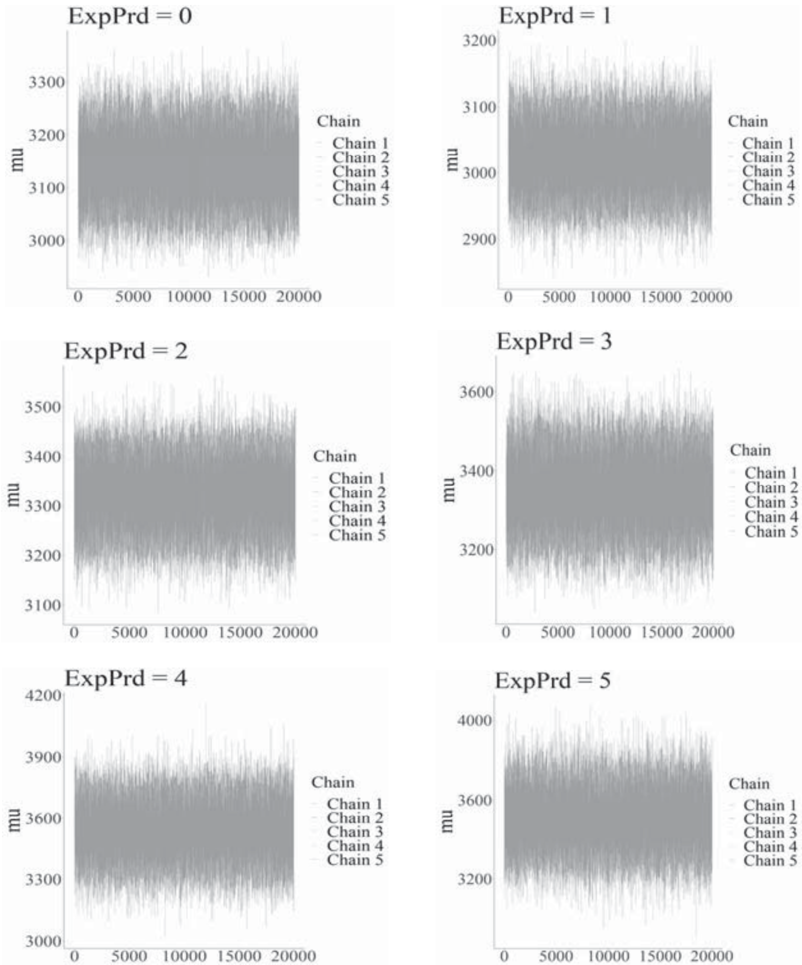


Figure 5

The Posterior Distributions of Vocabulary Size Estimates for Each Group at All Four Times (Marginal Means) Through MCMC Sampling with Medians and 95% CrIs

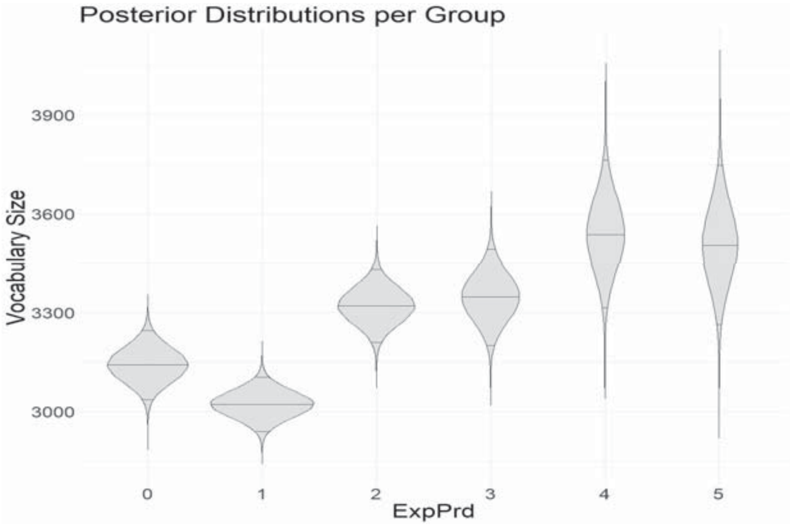


Table 3

Probabilities of Dominance Among the Marginal Means and Their Effect Sizes Estimated Through MCMC Sampling

Time	ExpPrd	Probability of dominance	Cohen's *d* EAP	Cohen's *d* post.sd	Cohen's *d* 2.5%	Cohen's *d* 97.5%
ALL	ExpPrd 0 > ExpPrd 1	.960	0.15	0.09	-0.02	0.33
	ExpPrd 0 < ExpPrd 2	.990	0.22	0.10	0.03	0.41
	ExpPrd 0 < ExpPrd 3	.988	0.26	0.12	0.03	0.49
	ExpPrd 0 < ExpPrd 4	.999	0.46	0.15	0.18	0.75
	ExpPrd 0 < ExpPrd 5	.996	0.46	0.17	0.13	0.79
	ExpPrd 1 < ExpPrd 2	1.000	0.38	0.09	0.21	0.56
	ExpPrd 2 < ExpPrd 3	.614	0.03	0.12	-0.19	0.26
	ExpPrd 3 < ExpPrd 4	.922	0.22	0.15	-0.08	0.52
	ExpPrd 4 > ExpPrd 5	.577	0.03	0.18	-0.31	0.38

Cohen's *d* [-0.02, 0.33] (little effect size). Hereafter, following Toyoda (2017), as for probabilities of dominance, an EAP (μ) of Cohen's *d*, a "post.sd" of Cohen's *d*, and a 95% CrI of Cohen's *d*, they are described as follows: ExpPrd 1 is lower than Non-EX (ExpPrd 0) with a probability of 96.0% and Cohen's $d = 0.15$ (0.09) [-0.02, 0.33] (little effect size). ExpPrd 2 is higher than Non-EX with a probability of 99.0% and $d = 0.22$ (0.10) [0.03, 0.41] (small effect size). ExpPrd 3 is higher than Non-EX with a probability of 98.8% and $d = 0.26$ (0.12) [0.03, 0.49] (small effect size). ExpPrd 4 is higher than Non-EX with a probability of 99.9% and $d = 0.46$ (0.15) [0.18, 0.75] (small effect size). ExpPrd 5 is higher than Non-EX with a probability of 99.6% and $d = 0.46$ (0.17) [0.13, 0.79] (small effect size). ExpPrd 2 is higher than ExpPrd 1 with a probability of 100% and $d = 0.38$ (0.09) [0.21, 0.56] (small effect size). ExpPrd 3 is higher than ExpPrd 2 with a probability of 61.4% and $d = 0.03$ (0.12) [-0.19, 0.26] (little effect size). ExpPrd 4 is higher than ExpPrd 3 with a probability of 92.2% and $d = 0.22$ (0.15) [-0.08, 0.52] (small effect size). ExpPrd 5 is lower than ExpPrd 4 with a probability of 57.7% and $d = 0.03$ (0.18) [-0.31, 0.38] (little effect size). In other words, when the period of EEE is less than two years (the group ExpPrd 1), there is an extremely small but negative effect with a probability of 96.0% compared to those who have no EEE (Non-EX). The group ExpPrd 1, which includes students whose period of EEE is less than two years, likely consists of students who discontinued their English learning experience shortly after they started due to various reasons such as lack of interest or motivation. If the students continued to be exposed with English early for two years or more, there are small effect sizes with a probability of 98.8% or more compared to those who did not experienced EEE (Non-EX). Also, the longer the students were exposed early, the more effective it was. However, this analysis is simple in that it ignores the "Time" variable and therefore also overlooks the hierarchy of data as mentioned in Section 2.5; it contains various issues related to the method of analysis. Consequently, we need to limit our interpretation here to general tendencies.

3.4 The Hierarchical Modeling Through Bayesian and ML Estimation

The intraclass correlation coefficient (ICC) was estimated as ρ = .0316 (ρ indicates *rho*). The ICC might seem to fall short of the lenient standard of .05 (the strict standard is 1.0) or more (Shimizu, 2014). However, since ICC is originally and theoretically only for the random-intercept model (Hox 2010), it should be considered merely for reference in this study where hierarchical modeling with both random intercepts and random slopes is adopted. Then, the design effect (DE) was calculated as DE = 8.94, indicating that the DE met criteria (standard is 2.0 or more [Shimizu, 2014]) as an index representing hierarchy. The individual line graphs and the Spaghetti graph of receptive vocabulary sizes of all 252 students over three years of high school are shown in Figures 6 and 7, respectively. Observing Figures 6 and 7 carefully, we can see that the intercepts of the 256 lines are random and the slopes are also random. This suggests that the longitudinal data of this study seem to fit the hierarchical modeling with both random intercepts and random slopes. Considering Figures 1, 6, and 7 comprehensively, the author judged that adopting the hierarchical modeling shown in Figure 2 for this research data is appropriate. The hierarchical modeling shown in Figure 2 was estimated through both Bayesian estimation and ML estimation.

As for the Bayesian estimation, the parameters were estimated by MCMC (HMC algorithm and NUTS) with normal distributions as prior distributions (a default setting of "stan lmer function" in the rstanarm package [Stan Development Team 2023a]) and are shown in Table 4-1. The indicators for the convergences of their estimates through MCMC sampling are shown in Table 4-2. The traceplots of MCMC sampling for the parameters of γ_{00} (described as "(Intercept)"), γ_{01} (described as "ExpPrd"), γ_{10} (described as "Time"), γ_{11} (described as "Time:ExpPrd"), and σ (described as "sigma") are displayed in Figures 8-1, 8-2, 8-3, 8-4, and 8-5, respectively (left side). Their corresponding posterior distributions, along with medians and 95% CrIs, are presented in Figures 9-1, 9-2, 9-3, 9-4, and 9-5, respectively (right side). The autocorrelations of MCMC sampling for these five parameters are depicted in Figures 10-1, 10-2, 10-3, 10-4, and 10-5, respectively. It

Figure 6

The Individual Line Graphs of Vocabulary Sizes of All Students Over Three Years of High School

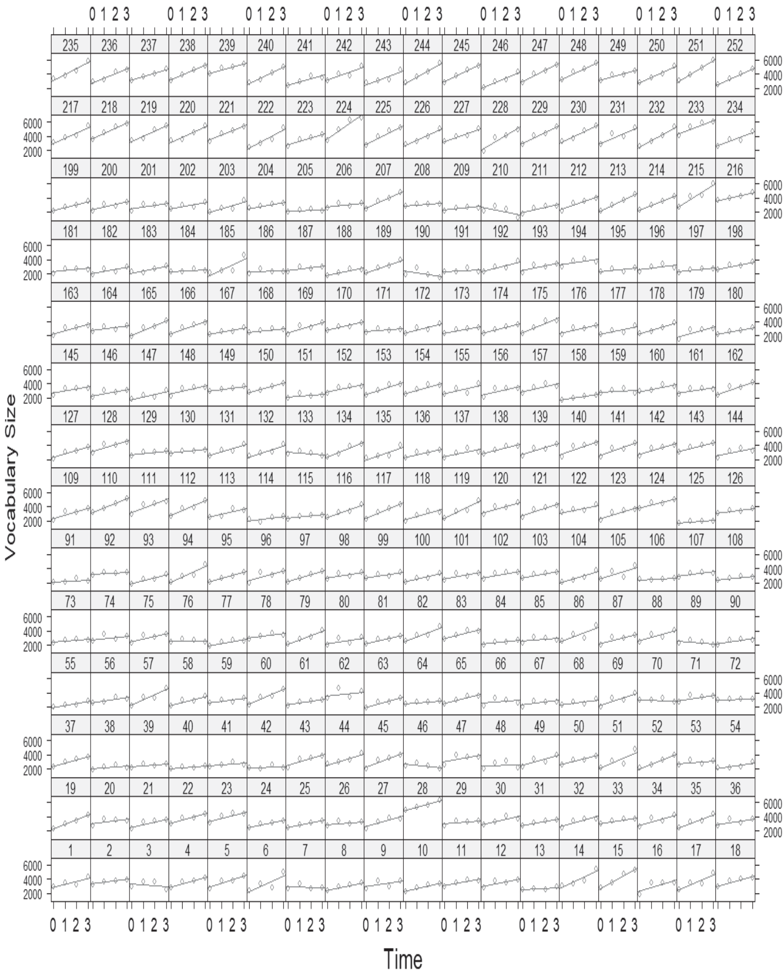


Figure 7

The Spaghetti Graph of Vocabulary Sizes of All Students Over Three Years of High School

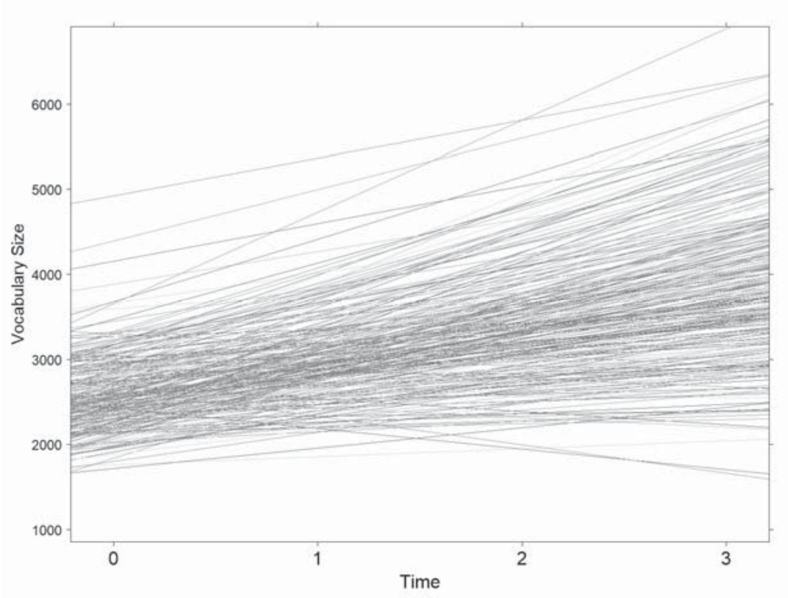


Table 4-1

Various Point Estimates of the Mathematical Model for the Bayesian Hierarchical Modeling Through MCMC Sampling

Estimator		**EAP (μ)**	EAP post.sd	95% CrI EAP 2.5%	95% CrI EAP 97.5%
The Hamiltonian Monte Carlo algorithm and the no-U-turn sampler	γ_{00}	2,511.2	42.1	2,429.2	2,593.5
	γ_{01}	68.0	19.4	29.8	105.8
	γ_{10}	354.1	23.3	308.4	399.8
	γ_{11}	24.5	10.8	3.4	45.6
	τ_{00}^2	$(327.5)^2$	$(131.7)^2$	$(275.8)^2$	$(79.5)^2$
	τ_{01}	40,892.4	6,733.8	27,734.1	54,169.9
	τ_{11}^2	$(186.4)^2$	$(72.7)^2$	$(158.7)^2$	$(214.1)^2$
	σ^2	$(334.9)^2$	$(3.23)^2$	$(17.8)^2$	$(18.9)^2$
	$r = \tau_{01}/(\tau_{00}*\tau_{11})$	.670			

Table 4-2

The Indicators for the Convergence of the Estimates for Table 4-1 Through MCMC Sampling

		n_{eff}	n_{eff}/N ratio	mcse	mcse/ post.sd	rhat
		as to EAP	as to EAP	as to EAP	as to EAP	as to EAP
The Hamiltonian Monte Carlo algorithm and the no-U-turn sampler	γ_{00}	86,719	0.867	0.143	0.003	1
	γ_{01}	90,745	0.907	0.064	0.003	1
	γ_{10}	99,685	0.997	0.074	0.003	1
	γ_{11}	106,968	1.070	0.033	0.003	1
	τ_{00}^2	27,639	0.276	104.4	0.006	1
	τ_{01}	22,247	0.222	45.1	0.007	1
	τ_{11}^2	17,170	0.172	40.4	0.001	1
	σ^2	18,718	0.187	0.08	0.000	1
	$r = \tau_{01}/(\tau_{00}*\tau_{11})$					

Table 5

Various Estimates of the Mathematical Model for the Hierarchical Modeling Through ML Estimation

Estimator		**Estimate**	*SE*	95% CI Lower	95% CI Upper	*df*	*t*	*p*
The Maximum Likelihood Estimation	γ_{00}	2,511.2	41.9	2429.1	2593.3	252	59.9	< .001
	γ_{01}	68.0	19.4	30.0	106.0	252	3.50	< .001
	γ_{10}	354.0	23.1	308.7	399.3	252	15.3	< .001
	γ_{11}	24.5	10.8	3.33	45.7	252	2.29	< .05
	τ_{00}^2	$(327.0)^2$						
	τ_{01}	40,914.2						
	τ_{11}^2	$(184.0)^2$						
	σ^2	$(334.2)^2$						
	$r = \tau_{01}/(\tau_{00}*\tau_{11})$	0.68						

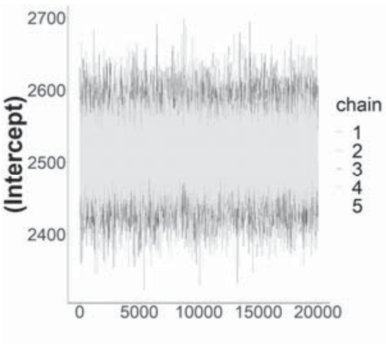
Note: AIC = 15,301.1, BIC = 15,340.4, logLik = -7642.5, deviance = 15,285.1, df.resid = 1,000

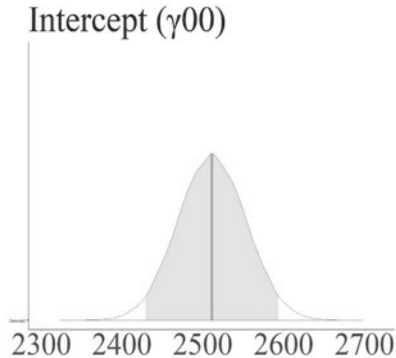
is observed that the respective autocorrelations quickly diminished to extremely small correlation coefficients following the commencement of MCMC sampling. No issues related to convergence in estimating the parameters of Bayesian hierarchical modeling through MCMC sampling were detected from (1) the respective indicators of n_{eff}/N ratios, mcse/post.sd, and rhats in Table 4-2, (2) the traceplots in Figures 8-1 through 8-5 , (3) the posterior distributions in Figures 9-1 through 9-5, and (4) the autocorrelations in Figures 10-1 through 10-5.

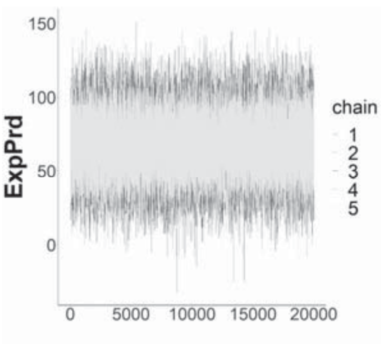
For the ML estimation, the parameters were estimated and are presented along with the model fit indicators (AIC = 15,301.1, BIC = 15,340.4, logLik = -7,642.5, deviance = 15,285.1, df.resid = 1,000) in Table 5.

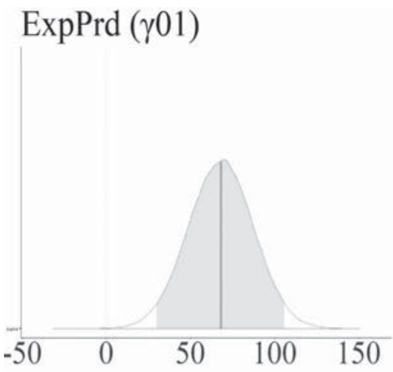
Figures 8-1, 8-2, 8-3, 8-4, and 8-5 (Left side) and Figures 9-1, 9-2, 9-3, 9-4, and 9-5 (Right side)

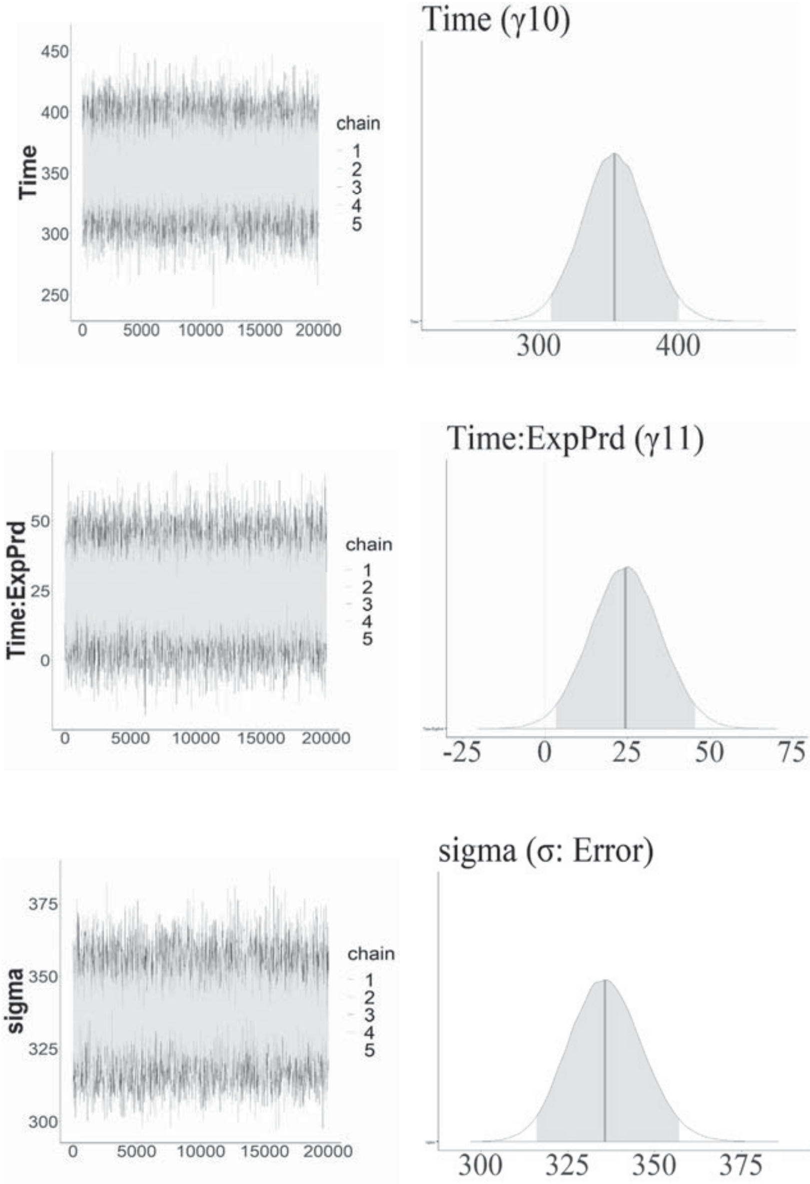
The Traceplots of MCMC Sampling for the Parameters of γ_{00} (Described as "(Intercept)"), γ_{01} (Described as "ExpPrd"), γ_{10} (Described as "Time"), γ_{11} (Described as "Time:ExpPrd"), and σ (Described as "sigma"), Respectively (Left Side); Their Respective Posterior Distributions with Medians and 95% CrIs (Right Side)










Time
450
400
350
300
250
0
5000
10000
15000
20000
chain
1
2
3
4
5
Time (γ10)
300
400
Time:ExpPrd
50
25
0
0
5000
10000
15000
20000
chain
1
2
3
4
5
Time:ExpPrd (γ11)
-25
0
25
50
75
sigma
375
350
325
300
0
5000
10000
15000
20000
chain
1
2
3
4
5
sigma (σ: Error)
300
325
350
375

Figures 10-1, 10-2, 10-3, 10-4, and 10-5

The Autocorrelations of MCMC Sampling in the Parameters of γ_{00} (Described as "(Intercept)"), γ_{01} (Described as "ExpPrd"), γ_{10} (Described as "Time"), γ_{11} (Described as "Time:ExpPrd"), and σ (Described as "sigma"), Respectively

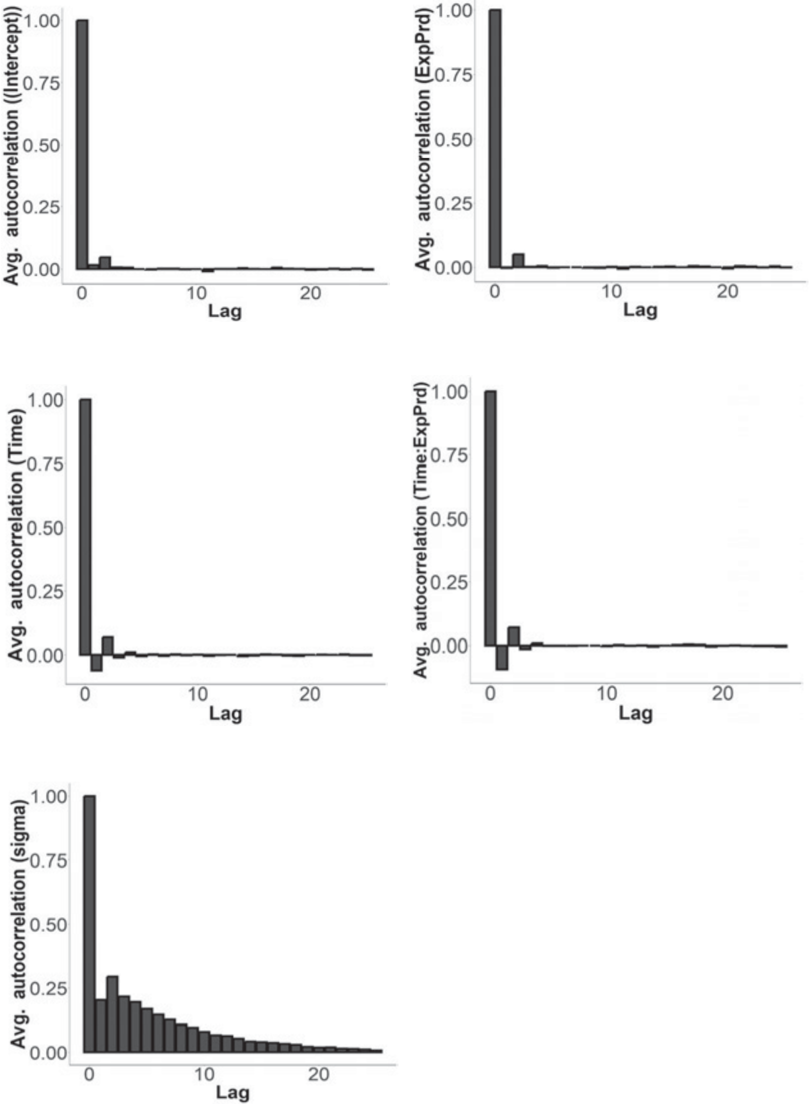


Figure 11-1

The Estimated Equations for the Hierarchical Modeling Through MCMC Sampling

Level 1 :

$$\text{Vocabulary}_{ij} = \beta_{0j} + \beta_{1j}\ \text{Time}_{ij} + \varepsilon_{ij} \tag{3.1}$$

$$\varepsilon_{ij} \sim N(0, [334.9]^2) \tag{3.2}$$

Level 2 :

$$\beta_{0j} = 2{,}511.2 + 68.0 \times \text{ExpPrd}_j + u_{0j} \tag{3.3}$$

95% CrI [2,429.2, 2593.5]; [29.8, 105.8]

$$\beta_{1j} = 354.1 + 24.5 \times \text{ExpPrd}_j + u_{1j} \tag{3.4}$$

[308.4, 399.8]; [3.4, 45.6]

$$(u_{0j}, u_{0j})' \sim MVN(\mathbf{0}, \boldsymbol{T}) \tag{3.5}$$

$$\boldsymbol{T} = \begin{bmatrix} (327.5)^2 & 40{,}892.4 \\ 40{,}892.4 & (186.4)^2 \end{bmatrix} \tag{3.6}$$

$$r = .670 \tag{3.7}$$

Figure 11-2

The Estimated Equations for the Hierarchical Modeling Through ML Estimation

Level 1 :

$$\text{Vocabulary}_{ij} = \beta_{0j} + \beta_{1j}\ \text{Time}_{ij} + \varepsilon_{ij} \tag{4.1}$$

$$\varepsilon_{ij} \sim N(0, [334.2]^2) \tag{4.2}$$

Level 2 :

$$\beta_{0j} = 2{,}511.2 + 68.0 \times \text{ExpPrd}_j + u_{0j} \tag{4.3}$$

95% CI [2429.1, 2593.3], $p < .001$; [30.0, 106.0], $p < .001$

$$\beta_{1j} = 354.0 + 24.5 \times \text{ExpPrd}_j + u_{1j} \tag{4.4}$$

[308.7, 399.3], $p < .001$; [3.33, 45.7], $p < .05$

$$(u_{0j}, u_{0j})' \sim MVN(\mathbf{0}, \boldsymbol{T}) \tag{4.5}$$

$$\boldsymbol{T} = \begin{bmatrix} (327.0)^2 & 40{,}914.2 \\ 40{,}914.2 & (184.0)^2 \end{bmatrix} \tag{4.6}$$

$$r = .68 \tag{4.7}$$

As a representative point, the results of the EAP estimates for each parameter are incorporated into the mathematical model depicted in Figure 2 and are illustrated in Figure 11-1 for MCMC estimation and Figure 11-2 for ML estimation. Focusing on γ_{01} in Equation (3-3) by MCMC estimation, it can be inferred that an increase of one in ExpPrd will result in an increase of 68.0 words upon entering high school (EAP = 68.0, post.sd = 19.4, 95% CrI [29.8, 105.8]). The same applies to γ_{01} in Equation (4-3) by ML estimation (estimate = 68.0, Standard Error [*SE*] = 19.4, 95% Confidence Interval [CI] [30.0, 106.0], $t(252) = 3.50$, $p < .001$). Focusing on γ_{11} in Equation (3-4) by MCMC estimation reveals that if ExpPrd increases by one—that is, if EEE experience increases by one in two-year intervals—there is a high probability that growth over one year of high school is slightly larger (24.5 words more) (EAP = 24.5, post.sd = 10.8, 95% CrI [3.4, 45.6]). The same applies to γ_{11} in Equation (4-4) by ML estimation (estimate = 24.5, *SE* = 10.8, 95% CI [3.33, 45.7], $t(252) = 2.29$, $p < .05$).

Considering Section 3.2 (Table 1-1 and Figure 3), Section 3.3 (Table 2-1, Table 3, and Figure 5), and this Section 3.4 (Table 4-1, Table 5, Figures 11-1 and 11-2) comprehensively, it is suggested that at the time of high school entrance (intercepts), EXs who experienced EEE for two years or more have higher estimated values for receptive vocabulary size than Non-EX, and that their growth (slopes) over three years of high school is slightly larger.

The *r* in Equation (1.7) is estimated as .670 through MCMC estimation and .68 through ML estimation, as shown in Tables 4-1 and 5, respectively. The *r* in Equation (1.7) represents the partial correlation between β_{0j} (intercept) and β_{1j} (slope), excluding the influence of ExpPrdj (Ozaki et al., 2019). It is observed that β_{1j} is positively and moderately correlated with β_{0j}. This implies that there is a moderate positive correlation between the receptive vocabulary size at the time of entering high school and the increase in receptive vocabulary size each year thereafter, once the influence by ExpPrdj is removed. This is also considered an academically interesting and significant finding.

Regarding the differences between Bayesian estimation through MCMC sampling and ML estimation, both methods estimated similar numerical values for each parameter, despite their fundamentally different philosophical paradigms. Bayesian estimation through MCMC is not yet widespread in L2 Teaching/Learning

Research, making this an academically interesting and important finding. The stability of the estimation in this study is confirmed.

3.5 Limitations

As previously described in Katagiri (2019), confounding factors (spurious correlations) may exist, potentially affecting both explanatory variables (independent variables) and objective variables (dependent variables). These confounding factors are thought to be related to parents' educational backgrounds, vocational backgrounds, income levels, among others. The influence of these confounding factors was not eliminated or reduced in this study.

4. Conclusion

For the first purpose of this study, it is indicated that at the time of high school entrance (intercepts), EXs who experienced EEE for two years or more have higher estimated values for receptive vocabulary size than Non-EX who experienced no EEE, and this trend continued throughout the three years of high school.

For the second purpose, it is indicated that EXs who experienced EEE for two years or more exhibit slightly larger growth (slopes) over three years of high school.

For the third purpose, it is indicated that both Bayesian estimation through MCMC (HMC algorithm and NUTS) sampling and ML estimation estimated similar numerical values for each parameter of the hierarchical modeling, despite their fundamentally different philosophical paradigms.

A moderate positive correlation (r = .670 [Bayesian estimation] or r = .68 [ML estimation]) was found between the receptive vocabulary size at the time of entering high school (intercept; β_{0j}) and the increase/growth in receptive vocabulary size each year (slope; β_{1j}), after removing the influence of early English exposures/education. This suggests that the more words the students know at the time of entrance, the more words they will learn during the three years of high school.

In summary, having been exposed/educated to English early in their childhood for two years or more was likely to confer an advantage both in their receptive vocabulary size at the time of high school entrance and in vocabulary learning during the three years of high school. Consequently, at the time of university entrance examinations, they were likely to have an advantage in their receptive English vocabulary size.

Notes

1. In the preparation of this manuscript, the author utilized Microsoft Bing AI to enhance the English sentences; however, it did not contribute to the generation of any original ideas.

References

Bates, D., Mächler, M. Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/ jss.v067.i01

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Haebara, T. (2014). *Zoku shinri toukeigaku no kiso* [Basics of psychological statistics (Continued from the last issue)]. Yuhikaku Publishing.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.

Katagiri, K. (2007). Examining the rationale of vocabulary size tests based on word-frequency levels. *The Japan Language Testing Association Journal*, *10*, 36-55. https://doi.org/10.20622/jltaj.10.0_36

Katagiri, K. (2009). A three-year longitudinal study of vocabulary size in Japanese SHS students and a description of their developmental patterns. *ARELE*, *20*, 131-140. https://doi.org/10.20581/arele.20.0_131

Katagiri, K. (2010). Three-year longitudinal study of the progress in listening and reading proficiency of Japanese senior high school students and their developmental patterns. *ARELE*, *21*, 221-230. https://doi.org/10.20581/arele.21.0_221

Katagiri, K. (2019). 1990-nendai ni okonawareta souki eigo kyouiku to sono kouka: Koukou 3-nenkan no juyou goi chishiki-ryou no menkara koukaryou to beizu de kyouiku kouka wo kenshou suru [Early English education in 1990s and its educational effects: Examining them with analyses of effect sizes and Bayes Factors through three-year longitudinal measurements of receptive vocabulary sizes in senior high school]. *Senshu Journal of Foreign Language Education*, *47*, 1-27. https://doi.org/10.34360/00006530

Katagiri, K., & Mochizuki, M. (2002). Goi saizu tesuto: Mittsuno han no heikousei [Mochizuki's (1998) Vocabulary Size Test: the parallelism of the three forms]. Paper presented at the 28th annual meeting of the Japan Society of English Language Education, held at Kobe University in Kobe on August 23, 2002.

Komori, M. (2022). *R to Stan de hajimeru shinrigaku no tameno jikeiretsu bunseki nyumon* [Starting with R and Stan: Introduction to time series analysis for psychology]. Kodansha.

Kubo, T. (2012). *Deta kaiseki no tameno toukei moderingu nyuumon* [Introduction to statistical modeling for data analysis]. Iwanami shoten.

Matsuishi, Y., Sakuramoto, H., Hoshino, H. et al. (2020). Down syndrome reduces the sedative effect of midazolam in pediatric cardiovascular surgical patients. *Scientific Reports*, *10*, 2148. https://doi.org/10.1038/s41598-020-58283-1

Ministry of Education, Culture, Sports, Science and Technology [MEXT]. (2017). *The course of study for elementary school (Released in March 2017)*. https://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm

Ozaki, Y., Kawabata, K., & Yamada, T. (Eds.). (2018). *R de Manabu maruchi reberu moderu (nyumon hen*) [Learning multilevel model with R (Introductory edition)]. Asakura Publishing.

Ozaki, Y., Kawabata, K., & Yamada, T. (Eds.). (2019). *R de Manabu maruchi reberu moderu (jissen hen*) [Learning multilevel model with R (Practical edition)].

Asakura Publishing.
Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079
R Core Team. (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* (4.3.1) [Computer software]. https://www.R-project.org/
RStudio Team. (2023). *RStudio: Integrated Development for R* (2023.09.0) [Computer software]. http://www.rstudio.com/
Shimizu, H. (2014). *Kojin to syuudan no maruchi reberu bunseki* [Multilevel analysis of individuals and groups.] Nakanishiya Publishing.
Stan Development Team. (2023a). *RStan: the R interface to Stan* (R package) (2.26.23) [Computer software]. https://mc-stan.org/
Stan Development Team. (2023b). *Stan Reference Manual Version 2.33.* https://mc-stan.org/users/documentation/
Suzuki, Y., Tanaka, N., & Akiyama, H. (2020). Attempt of Bayesian estimation from left-censored data using the Markov chain Monte Carlo method: Exploring Cr(VI) concentrations in mineral water products. *Food Safety*, *8*(4), 67–89. doi: 10.14252/foodsafetyfscj.D-20-00007
Toyoda, H. (2017). *Shintei―Shinri toukei hou* [Psychological statistics (New edition)]. Housou daigaku kyouiku shinkoukai.