O DOCUMENTING THE DIVERSITY OF LEARNER ACHIEVEMENTS IN ASIAN LANGUAGES USING COMMON MEASURES

Catherine Elder, Hyejeong Kim & Ute Knoch University of Melbourne caelder@unimelb.edu.au

While there have been a number of studies exploring the impact of time-on-task and language background on language achievement for both English and other languages, the Student Achievement in Asian Languages Education (SAALE) project constitutes the first systematic attempt to gather empirical evidence of these effects for four Asian languages (Chinese, Indonesian, Japanese and Korean) in Australian schools. The paper focuses on the approach adopted to measure and exemplify the diverse nature of learner achievements in the context of concern. This approach involved a) the gathering of information about the language background and prior learning experience of the study's participants in order to establish learner sub-groups for subsequent analysis, b) the use of common assessment procedures at each level of schooling to compare levels of achievement across learner sub-groups and c) the analysis of samples of each sub-group's performance by teams of teacher experts to develop rich descriptions of achievement reflecting the different dimensions of diversity relevant to each language.

The paper outlines the methodology adopted for the study, and reports briefly on the overall findings, Particular attention is paid to the challenges encountered in undertaking the research and to the further efforts that are needed to build on the project outcomes.

KEY WORDS: Learning Asian languages, language background, time-on-task

INTRODUCTION

Scarino (this issue) has offered a rationale for a context-sensitive approach to developing language descriptions that goes beyond the generalized approach adopted in the standards-based frameworks that currently dominate the field of languages education in Australia and elsewhere. Critical dimensions of diversity in the Australian language learning context are the institutional time allocated to language learning on the one hand and the language background of the learner on the other. This paper offers a brief historical overview of the literature highlighting the importance of these variables for language learning. Then, by way of background for the papers which follow, it sketches the methodology adopted in the SAALE project to a) explore the influence of time-on-task and language background variables on what learners achieve in the four Asian languages (Chinese, Indonesian, Japanese and Korean) targeted for the current study and to b) produce detailed and empirically grounded descriptions which are sensitive to the influence of these variables on

achievement. It then gives a brief overview of the study's findings, which will be fleshed out in the language-specific papers that follow. The paper concludes with a reflection on the challenges faced in conducting the study, and on possible avenues for further research.

LITERATURE REVIEW

Questions about the time required for effective language learning have a long history, dating back to the 1960s with investigations of how much time was needed to obtain particular levels of proficiency in different languages (Cleveland, Mangone & Adams, 1960) and of the optimal time at which foreign or second language instruction should begin (Carroll, 1967; Stern, 1985). Both Carroll (1967) and Stern (1985) showed a link between the amount of instructional time and proficiency, inviting the conclusion that the more time is provided, and the earlier language learning starts, the better the outcomes will be. Given that the time allocated to language learning within the Australian mainstream school curriculum is often very limited (e.g. see Victoria, Department of Education and Early Childhood Development [DEECD], 2010), it should be expected that the rate of learning will be relatively slow. This has been one of the justifications for efforts in Australia and elsewhere to increase the duration of instruction by introducing language study in the primary school (Pufahl, Rhodes & Christian, 2000). A second assumption underpinning early introduction of language study is that young children are better language learners than older ones (Blondin et al., 1998; Vinjé, 1993; Clyne, Jenkins, Chen, Tsolalidou & Wallner, 1995). However, studies that have compared early and later starters in Australia and other countries have produced equivocal findings (Brown, Hill & Iwashita, 2000; Burstall, 1975; Hill, Davies, Oldfield & Watson, 1997). It seems that, whether for Asian or other languages, the benefits of primary school instruction on subsequent language learning achievements are sometimes short-lived, perhaps because early learning achievements are neither recognized nor built upon by teachers due to poor continuity between primary and secondary school programs (see Hill, 2012; Johnstone, 2000, 2006; Oostdam& Van Toorenburg, 2002). Given these uncertain benefits, the link between an early start and subsequent learning achievements is explored in the SAALE study by Kohler (this issue) in relation to Indonesian, which is one of the more widely taught languages in Australian primary schools.

While early start programs potentially increase the duration of learning and hence the ultimate level of achievement, intensity of learning may also play a role. A number of studies in the late 1980s and 1990s attempted to explore this issue experimentally, finding advantages for massed over distributed instruction for groups receiving similar amounts of overall instruction but spread over different periods (Collins, Halter, Lightbown & Spada, 1991; Lapkin, Hart & Swain, 1995; Spada & Lightbown, 1989). Studies of language gains from immersion contexts in Australia generally support the notion that learning languages more intensely is beneficial (e.g. see Clyne, 1986; de Courcy, 2002; Elder, 1989; Farmer,

2006; Lorch, McNamara & Eisikovits, 1992; Molyneux, 2004) although such studies have not involved any direct comparison with non immersion classrooms, as was possible with the Japanese primary school component of the SAALE project where both immersion and non immersion learners were included in the sample (Scarino et al., 2011). In any case it is unclear whether it is simply intensity of instruction that contributes to positive learning outcomes or other factors such as motivation or the focus of instruction in such programs which tends to be on content rather than simply on language learning for its own sake.

Study abroad has likewise been found to accelerate progress in various languages (e.g. see DeKeyser, 1991; Freed, 1995; Marriott, 1993; Wilkinson, 1998), although again the findings are mixed given the multiple factors involved. More recent research has paid more attention to these factors noting that not all study abroad students take full advantage of opportunities for target language exposure (Serrano, Tragant & Llanes, 2012) and that intensity of exposure in the study abroad context is no guarantee of the quality of the learning experience (Dufon & Churchill, 2006).

It is worth noting that the above studies deal with a range of target languages and one of the reasons for the mixed findings may be not only the contexts of learning but also the nature of the languages themselves, some of which are regarded as more challenging to learn than others. This relative difficulty notion has its roots in contrastive analysis and in particular the work of Lado (1957), who attempted systematic comparisons of different languages with a view to identifying the elements likely to cause difficulty or 'interference' in learning a new language and hence slow down the process of forming new habits on which this learning was allegedly based.

Despite of a changed view of learning as a creative construction process governed by the constraints of universal grammar, the study of cross-linguistic influence (CLI) continued to be influential in second language acquisition research. The emphasis however shifted from interference resulting from points of difference between languages to the facilitative effects of similarity. Odlin (1989), for example, claimed that similarity between languages could confer important advantages and cited the estimates from the US Foreign Service Institute (FSI) regarding the instructional time needed to attain elementary proficiency in different languages (http://www.govtilr.org/). The estimates, based on decades of experience of teaching foreign languages intensively to adults in military contexts, indicate that even basic achievements in Chinese, Japanese and Korean will take longer for native English speaking learners than is the case for Indonesian, which may however be more time consuming than a Latinate language like French.

One of the problems with the FSI estimates is that they assume English as the point of departure for the foreign language learner. This does not always apply in multilingual societies like Australia (of for that matter the United States) where the population of learners studying languages in the school contexts is increasingly heterogeneous. The task of

learning, as Corder (1981) noted, will be very different for learners with different first language backgrounds or prior experiences of learning other languages, whether formally or in naturalistic contexts.

Current research is more mindful of the fact that cross linguistic influences may derive not only from the first language but also from additional languages in the learner's repertoire, that transfer can work both positively and negatively at different levels of the language system and that it may be triggered by multiple factors which interact with one another in complex ways (Jarvis & Paylenko, 2008). There is also growing recognition of fact that first language background or prior experience of learning additional languages may be a surrogate for factors such as familiarity with the target language culture and learning style (Abbott, 2006; Rao, 2001) which also play a role in language learning. Indeed, the broader cultural and educational factors associated with language background may in some cases be just as influential as linguistic factors in determining the nature and level of achievement in school language learning contexts. This possibility is considered by Iwashita (this issue) in her investigation of levels of Japanese achievement among learners from diverse language backgrounds including those languages (such as Korean and Chinese) that are "related" in the sense of sharing some common features with Japanese. The presence of sizeable numbers of such learners in Japanese primary and secondary school classrooms and within the SAALE sample affords a unique opportunity to explore the potential link between L1 background and level of learning achievement.

While it appears that Japanese study is an attractive option for learners of Korean and Chinese background in Australian schools, these learners may also, or instead, choose to study their mother tongue in the Australian school context. Policy rhetoric surrounding the teaching of Asian languages in Australian schools has tended, at least until recently, to target students from English-speaking backgrounds (e.g. Rudd, 1994). Nevertheless, since immigrants from Asia constitute one of the fastest growing community language groups in Australia, their children now constitute a sizeable proportion of those studying their home or community language within the mainstream school system (Slaughter, 2007a). In addition, among those enrolled for languages subjects at the senior secondary level are growing numbers of international students who have been educated in countries where the target language is the official medium of communication.

The presence of these "background learners" in language classrooms alongside "foreign language learners" whose only exposure to the target language is in the classroom has long been the subject of contention particularly at senior secondary level, where background and non-background learners compete for places in higher education (Elder, 1997; Ozolins, 1993). However, in spite of the salience of the language background issue in public discourse, it has been the subject of relatively little research in Australia.

Two notable exceptions are studies by Clyne, Fernandez, Chen and Summo-O'Connell (1997) and Elder (1996, 1997, 2000a, 2000b) conducted during the 1990s. These studies demonstrated the complexity of categorizing the heterogeneous population of learners in Victorian language classrooms given learners' varying degrees of linguistic and cultural association with the target language via one or both of their parents or grandparents. Clyne et al. (1997) examined the linguistic performance of a small sample of background learners of Chinese and of two European languages (Italian and German) who were studying their home or ancestral language in Years 10, 11, and 12 of secondary school. They documented particular linguistic features of their performance requiring pedagogical intervention including instances of lexical, semantic and syntactic transference from both English and from non-standard dialects and highlighted the importance of catering for the particular needs of background learners in the interests of maintaining and developing their existing language skills.

Elder (1996, 1997, 2000a, 2000b) conducted large-scale comparisons of background and nonbackground learners of languages in Australian schools based on their performance on common language examinations. Her focus was on junior and senior secondary school learners of Chinese, Italian, and Modern Greek. She found that the achievement gap between background and non-background learners as defined by self-report data about language background was, on average, far greater for Chinese than for the other two languages. This was attributed to the fact that many Chinese learners were recent arrivals in Australia with active native-like competence acquired through schooling in their home country (whereas the Greek and Italian learners tended to be second or third generation immigrants with far more limited exposure to their home language). There was nevertheless huge variation in achievement within the background learner group for each language according to both the nature of the task and a range of factors including the variety and status of the language to which the individual students had been exposed (i.e. how closely it conformed to the taught language of schooling), the student's age at immigration (if born in a country where the target language was spoken), the first language of the parents and the extent of prior schooling or home literacy in the target language. Elder drew attention to the challenges this diversity of background posed for language assessment, pointing to the problems for test validity of using common examinations to assess learners with markedly different ability profiles. She called for clearer thinking at the policy level about the particular needs of background and non-background learners as well as a rethink of the reductionist foreign language learner/native speaker dichotomy which characterized much of SLA research at the time.

Different states in Australia have attempted to address the diversity of background within the learner population by allocating learners to different assessment levels or streams at least in the upper secondary school, based on learners' self-reported prior learning and/or exposure to the target language. However the nomenclature for these different levels and the criteria for grouping

learners differ from state to state (Scarino et al., 2011; Slaughter, 2007b) and are generally based on expert opinion rather than on empirical evidence of differences in achievement.

New insights on background learner issues have emerged from research on what are known as 'heritage languages' in Canada and the US. Indeed, heritage language education is now seen as an academic field in its own right. As in Australia, there is some debate around who qualifies for heritage language learner status (Carreira, 2004; Valdés, 2001; Wiley, 2001) with arguments around whether group membership should be based on self-ascribed identity, family background or language proficiency. In the literature dealing specifically with heritage language pedagogy and assessment proficiency-based definitions are more common (Llosa, in press), although there is evidence (Kagan, 2005) that self-reported data on language background and home language use has value in predicting learner proficiency, as was also found in the SAALE project.

In the US, by far the largest group of heritage learners are Latinos, but other language communities including Korean and Chinese are also receiving increasing attention and have been the subject of published research (treated in more detail in the final two papers in this issue by Kim and Scrimgeour). Federal language policy initiatives have given greater explicit attention to these learners than is the case in Australia, particularly since 9/11, when it was recognized that the nation's existing linguistic resources needed to be harnessed in the interests of national security (King & Ennser-Kananen, 2012). In addition, many US university programs offer specialist study tracks for heritage languages learners (Kondo-Brown, 2008) in contrast to Australia, where the distinction between different types of learner is seldom explicitly drawn.

One of the more promising directions of US heritage language research is the systematic documentation of particular characteristics of the spoken and or written production of heritage language learners with regard to such areas as vocabulary (Polinksy & Kagan, 2007) tense aspect and modality (Montrul, 2002; Montrul & Perpiñán, 2011; Polinksy, 2008,) relative clauses (Kim, 2005) and reflexive pronouns (Kim, Montrul & Yoon, 2009), although this documentation tends to focus on particular features rather than on the multidimensional descriptions generated by the SAALE project. Also important are efforts to theorize the process of heritage language acquisition in ways that differentiate it from either first or second language learning (e.g. see Bolger & Zapata, 2011; O'Grady, Kwak, Lee & Lee, 2011) and to explain why acquisition amongst heritage language learners may be only partial and fall-short of nativelevel monolingual norms (O'Grady, Lee & Lee, 2011). Finally there appears to be growing interest in the pedagogical implications of heritage language research findings both for course placement and curriculum design (Carreira & Potowski, 2011). However, thus far little attention has been paid to making evidence-based research findings available in a form which is accessible to language teachers – a gap which the papers in this volume and the descriptions generated from the SAALE project are designed to at least partially address.

The above (admittedly selective) overview of the literature serves to buttress the assumptions driving the SAALE project, the basis for the sampling and grouping learners and for the descriptions of achievement that this project has generated. On the one hand it is hypothesized that that studying a new or additional language for a sustained period of time or intensively rather than in a "drip feed" fashion may result in higher levels of achievement. It is also posited that learning achievements in the target language will be influenced by a learner's first language as well as by any other additional language learning experiences or cultural exposure that s/he may have had. Similarly, it is proposed that heritage or background learners with some degree of home or community exposure and or schooling in the target language (or a variety of it) will differ in the level and also in the quality of their achievement from learners without any connection to the target language outside the classroom and indeed from native speakers in the country of origin whose acquisition can be considered complete and less vulnerable to attrition. Although these hypotheses require qualification, given individual variation and the potentially intervening variables alluded in the above review, they are worthy of exploration with the languages investigated in the current study for the reasons outlined by Scarino (this issue). The methodology for both quantitative and qualitative (descriptive) dimensions of the SAALE project will be outlined in the following section.

SAALE PROJECT METHODOLOGY

The methodology for the SAALE project involved the collection of data from learners in Australian schools via two questionnaires and an outcome measure. These instruments will be described in further detail below.

INSTRUMENTS

To gather information from learners about their language background and the time they had spent learning the language, two questionnaires were designed, the first eliciting information from participating schools about their programs, including institutional time-on-task variables (i.e. frequency, duration and intensity of language instruction within the program) and the language background profile of the enrolled students. The second questionnaire was designed for individual students, with similar questions about time-on-task, including whether they were continuing language study from primary school, and more detailed questions about the aspects of their language background (e.g. parents' first language, country of birth, years of instruction through the medium of the target language) considered likely to have a bearing on their level of performance. These combined sources of information gave us what we needed to operationalize the time-on-task and language background variables in our study and to link them to achievement data from each learner.

The outcome measure for the study was language learning achievement as evidenced on language tests/examinations. The particular challenge for the study was to devise a common measure for each year level that would allow us to compare the achievements of learners from different language backgrounds and with different amounts of instructional exposure. These tests therefore needed to span a broad range of abilities. At Year 12 we used existing examination data provided by the curriculum and assessment authorities in the participating states but reassessed learner performance against a common set of criteria devised expressly for the current study (see further details below). At Years 6/7 & 10 we used custom-built measures developed by language-specific panels comprising expert teachers convened for this purpose. The specifications for these tests were based loosely on the key performance measures (KPMs) developed from an analysis of curricula in each Australian state and territory. At Year 10 we developed common tests of Reading, Writing and Oral interaction (listening and speaking). At Year 6/7 the tests included Reading and Writing components only, because of the sensitivities associated with administering one-on-one oral tests for young learners who might be intimidated by having to perform solo in front of a stranger while at the very early stages of their language acquisition. Instead, a classroom observation protocol was developed to gain insights on oral interaction (including listening) skills during classroom visits by members of each language-specific panel. All custom-built tests were piloted on a small group of learners for each language and revised as required.

PARTICIPANT SAMPLE

The sample for the study was to be 400 students per level for each language and broadly representative of the distribution of language learners across State, Catholic and Independent schools in various Australian states. For pragmatic reasons we included only four of the seven Australian states for our sample (New South Wales, South Australia, Tasmania and Victoria) and sought cooperation from the relevant jurisdictions to assist us with selecting appropriate schools according to specified criteria, in order to capture the full range of relevant time and language background variables for our study. Because the data reported in subsequent papers in this issue relates only to Year 6/7 and Year 10 learners, only these two cohorts will be considered from here on. Further information about the Year 12 learners can be found in the full report on the project (Scarino et al., 2011).

The number of willing recruits at these lower year levels turned out to be far more limited than originally anticipated, Table 1 below shows the indicative numbers provided by the schools agreeing to participate (Column 1) and the actual numbers of students (Column 2) from whom we received data at the two year levels. More details on the profile of participating schools are provided in Scarino et al. (2011) and in the particular papers in this issue. It can be seen that the indicative numbers were close to or beyond our target for Year 6/7 Chinese, Indonesian and Chinese but well below target at Year 10, with the exception of Japanese. The actual numbers are far lower, for reasons which will be outlined below.

Table 1: Summary of student numbers in the sample

Language	Level	Indicative N	Actual N	
Chinese	Year 6/7	542	259	
	Year 10	282	133	
Indonesian	Year 6/7	372	131	
	Year 10	154	79	
Japanese	Year 6/7	418	173	
	Year 10	386	141	
Korean	Year 6/7	55	78	
	Year 10	29	20	

DATA COLLECTION PROCEDURES

Ethics clearance was sought and granted by the relevant jurisdictions and participating schools in each state. The schools in turn sought consent from parents and students in their language classes.

At both years 6/7 and 10, the Reading and Writing tests and learner background questionnaires were administered to students by the language teacher in each school over two lesson periods according to an invigilation protocol. The teacher also completed the program profile questionnaire. All material was then returned by post to the Language Testing Research Centre in Melbourne for processing. Oral tests were conducted by members of the language-specific panels convened for the project. At Year 10 a one-on-one interview was conducted with learners at a subset of schools. (Gathering oral data at all schools was not logistically feasible.) All interviews were digitally recorded for later rating and analysis. At Year 6/7 the classroom teacher was briefed on appropriate interactive activities to conduct with the class according to a set protocol. The interaction was digitally recorded by the panel chair, who also conducted (and recorded) a small group interview with 6-8 students from the class.

MARKING PROCEDURES

After the test papers for each language had been returned by the schools they were assigned for marking. Marking of the Reading tests followed a custom-built marking guide for each language indicating the range of acceptable responses for each item and the basis for scoring. Writing and Oral assessments were assessed against an agreed set of criteria that were the same across languages and year levels. For Writing, five categories were used to judge students' writing performance at each year level: content, vocabulary, forms/structures, discourse, and scripts and/or characters (depending on the language). The first four were used for all languages but the fifth one was omitted for Indonesian, being the only language of the four with a Romanised script. For the Oral assessment (listening and speaking combined),

seven rating categories were considered: content, vocabulary, forms/structures, fluency, intelligibility, comprehension, and discourse (see Scarino et. al., 2011, for the full set of criteria for each language and the descriptors developed for rating at each year level). (Oral interaction data gathered for Year 6 was not scored, but instead analysed qualitatively according to procedures described below.) The chair of each language panel organized training of raters using selected student samples. All samples were then double marked.

QUANTITATIVE ANALYSIS

The quantitative analysis of score data was for the purpose of exploring the links between language background and time-on-task variables. The results of this analysis also informed the learner groupings that were used for the subsequent qualitative process of developing descriptions of achievement. The steps involved in preparing for and undertaking the quantitative analysis will be outlined below. The process of developing descriptors will be described in a separate section.

Reliability analysis

Checks were first undertaken using the Cronbach's alpha statistic to ascertain the internal consistency of the Reading tests. Various inter-rater agreement statistics were calculated for the scores assigned for Writing and Oral tests. These estimates (reported in detail in Section 4.2 of Scarino et al., 2011) confirmed that for all languages the scores yielded by the instruments were sufficiently robust to be used for the subsequent analyses.

Selection and grouping of variables for analysis

The quantitative analysis for both Year 6/7 and Year 10 involved cross-referencing data from the language background questionnaires and school program profiles to scores on the Reading, Writing and Oral tests administered at both year levels.

Variables considered critical for this analysis, given the focus of the study, were the following:

a) Time-on-task, as indicated by years of community schooling, years of studying the target language at school, and mean hours/minutes per week of study (school program profile). Experience of studying another language (other than the target language) was also treated as a time factor given what was suggested in our literature review above, namely that the experience of studying any language can yield benefits for learning a new language. In the case of Japanese, we explored a further variable that applied to some of the primary school learners, that is, experience of bilingual education. For both Chinese and Japanese we were able to investigate in-country study experience given that a number of schools in the sample offered in-country programs.

b) First language background, as indicated by the birthplace of the learner and his/her parents and the L1 declared by the learner, as well as age of arrival in Australia and length of residence. The country of birth, and parent country of origin variables were grouped according to countries where the target language or a variety of it, was an official medium of communication/instruction, countries where another Asian language, which shared some common features with the target language, was the official medium, and countries where English or another language was the official medium.

Given the need to produce manageable groupings for further analysis, the project team and associated panels of expert language teachers came up with definitions of three major categories of learner language background: *first language learner*, *background language learner* and *second language learner*, as set out below. The labels assigned to each grouping are not intended to reflect those currently used by the various educational systems around Australia since, as already noted in the literature review, the meaning of these labels varies from state to state. The criteria for assigning learners to groups were as follows:

First language learner

Born in a country where the target language (or a variety of it) is an official medium *AND* arrived in Australia at the age of 8 or more (and therefore is likely to have had more than 2 years' experience of formal schooling through the medium of the target language, or related variety, before arrival in Australia) *AND* the target language was the first language used before starting school *AND/OR* uses the target language at home.

Background language learner

Born in Australia but with one or more parents born in a country where the target language is an official medium *AND* the target language was the first language used before starting school AND/OR uses the target language at home OR Born in a country where the target language (or a variety of it) is an official medium *BUT* arrived in Australia before the age of 8 (and therefore has limited experience of target language medium instruction).

Second language learner

Born in a country where the target language (or a variety of it) is *NOT* the official medium *AND* does not have more than one parent born in the country where the target language is official medium *AND first* language before school *AND/OR* language used at home is *NOT* the target language or a variety of it.

The criteria for these broad language groupings can be considered hypotheses about likely levels of achievement derived from our collective experience as teachers and/or researchers

informed by our readings from the research literature. Thus the *first language learner* is a recent arrival in Australia schooled in the country of origin through the medium of the target language and is likely to have age-appropriate literacy skills in that language far beyond what would be expected of those studying for a limited amount of time in second language classes in a country where the language is not the official medium. The *background language learner*, because of his/her parentage and exposure to the target language from an early age, is likely to have a strong linguistic and cultural association with the target language and to have acquired some degree of target language competence, whether active or passive, outside the school context, but less developed than that of the first language learner as defined above. The *second language learner*, who in most cases has had no sustained exposure to the target language or culture outside the classroom, is predicted to achieve at a lower level in the target language than members of the previous two learner groups.

Participants were assigned to one of the broad language background groupings based on their responses to the questionnaire. The validity of these groupings as predictors of achievement was explored by comparing mean scores on the tests for each group using inferential statistics (T-test or ANOVA and post-hoc comparison) where numbers permitted.

RESULTS

Results of the quantitative analysis will be reported only briefly here due to space constraints (see Scarino et al., 2011, for a full report).

Time-on-task factors were considered only for *second language learners* to avoid the confounding effects of language background, which for some learners affords opportunities for exposure far greater than those available in the context of school instruction. The effects of time-on-task factors (including duration of study and intensity of study and experience studying an additional language) proved to be inconsistent both across languages and year levels. This may be partly explained by insufficient variation within the sample, by the conditions of learning for certain languages and by the interaction between time-on-task and other variables. For example, primary school learners studying Japanese in schools offering partial immersion far outperformed those studying under limited exposure conditions, but such programs either do not exist, or were not available for scrutiny in the other Asian languages. Likewise, the effects of an early start were discernible for Chinese and Indonesian but not for the other languages. It is difficult to pin down the precise reasons for this inconsistency due to the host of other variables that could potentially interact with time-on-task (see further discussion in Kohler, this issue, in relation to Indonesian).

As for language background, our predictions about the differences in level of achievement according to language background were sustained in most cases, although the considerable variation with the *background language learner* groups for Chinese and Korean in particular

yielded test scores that were not always significantly lower than those of the *first language learners*. For Japanese, the L1 background of the learner (whether English, or Chinese and Korean) was found to make a difference to levels of achievement at both year levels (see Iwashita, this issue, for further information).

QUALITATIVE ANALYSIS

Drawing on the statistical analyses described above, a qualitative analysis was undertaken of student scripts and oral performances. These analyses involved all members of the language-specific panels, many of whom had also been involved in the marking process and were therefore thoroughly familiar with the assessment instruments, marking guides, and processes.

Determination of language background groupings for description

The language background groupings relevant to each particular language were determined by each panel according to the statistical results. While the criteria for *first language learner*, background language learner and second language learner remained as described on page 261, these groups were not retained for all languages if a) the groups did not perform at different levels as revealed by the statistical analysis and/or b) there were insufficient numbers in the group. For example, the partial immersion students at Year 6/7 in Japanese, as already noted, were both statistically distinct from other Year 6/7 Japanese learners and made up circa 30% of the sample. They were therefore treated as a separate group for the purposes of describing achievement. By contrast, there were insufficient numbers of first and background learners of Japanese at this year level to create three separate language background groups. The first and background language learners were therefore combined to form a single group.

Selecting samples for annotation

Selection of sample scripts for annotation took place at a 3-day intensive workshop held by each language panel. Before the workshop, writing test papers and recorded speech samples for each identified language subgroup had been ordered by test score to allow consideration of achievement levels within each group.

Within each subgroup, panel members, working collectively, selected 8-10 representative samples of students' performance at 'high' and 'average' score levels for closer qualitative analyses. This was done first for Years 6/7 and then for Year 10 students. The dividing line between high and average scorers depended on the score distribution for the test in question. It was not fruitful to consider low-scoring performances as this often entailed a short or indeed absent response. The selected samples of performance in the 'high' and 'average' range formed the basis for two sets of descriptors within each language background

subgroup, one representing the norm and the other representing what was achievable by higher performers from the relevant grouping.

Developing descriptors

To develop the descriptors for each language skill, each panel member worked mainly with the identified scripts referring from time to time to other examples in the full data set to ensure that justice was done to the range of achievements.

These annotations included:

- general observations on the characteristics of the 'high' and 'average' samples in relation to each of the assessment criteria for the particular assessment task;
- particular examples of the features of interest within each criterion, making sure that the annotations included sufficient detail to be meaningful;
- any features that exemplified the performance of those who had been shown, from
 the statistical analysis, to be more likely to perform at higher levels (e.g. in the case
 of Japanese, features characteristic of Korean and Chinese learners were included in
 the description of the high achieving group [see Iwashita, this issue]).

Drawing on the annotations made in the workshop the panel chairs drafted initial descriptions of achievement for the groups relevant to their particular language. These descriptions then went through several rounds of review, with feedback from other panel members, the whole research team, a wider group of experienced teachers and language experts drawing on their more generalized experience of student achievement and, finally, nominated representatives from the State, Catholic and Independent school jurisdictions. The final descriptions can be found at http://www.saale.unisa.edu.au/project.html.

Selection and annotation of exemplar tasks

Finally, to supplement the group descriptors at high and average levels of ability for each language background grouping, panel members selected sample scripts to exemplify particular trends in the data. For example, for Chinese as a second language at Year 6/7, the performance of a learner who had studied continuously from early primary school was included as one of the high achiever exemplars, given the statistical finding that such learners outperformed those who had studied for a more limited amount of time. Thus the exemplars, duly annotated and contextualized with relevant information about the learner's particular language background and study experience, served to highlight the observed statistical links between particular time-on-task variables and school achievement in the target language.

DISCUSSION

Thus far the process of exploring the effects of language background and time-on-task on school achievement in four Asian languages has been outlined, together with a summary account of how both the statistical results and the data which yielded these results was harnessed to develop context-sensitive descriptions of learner achievement at different levels of schooling. This process, while rigorously planned and executed, has a number of limitations that will now be discussed

First is the problem of relying on student self-report data for our language background groupings. There were inconsistencies in the way the questionnaires were completed suggesting that, in some cases, the learners had not understood what was expected for some of the questions or indeed may have not wished to own to having a background in the target language (see Elder, 2000a, for a discussion of the potential consequences of self-declaration). This means that the grouping of learners according to various dimensions of background and study experience may not be fully accurate.

More important is the limitation of sample size, which, as noted above, was far smaller than anticipated. The sample moreover was not fully representative of the state of play for languages in the relevant states, since the recruitment process captured a disproportionate number of independent schools with established programs where we might expect learners to achieve at relatively higher levels. Our limited success in recruiting students for this project was due to a number of factors, including the timing of the administration towards the end of the school year, a general reluctance by school administrators and/or teachers to engage in a research activity which encroached on class time and required teacher supervision and consent from each participating student and their parent/s. Diffidence by teachers about learner performance on our tests and how the results of these tests might be used may also, we suspect, have been a factor. In spite of our attempts to explain the purpose of the project, not all teachers understood that is was essentially a research activity rather than a program evaluation exercise.

The teachers' diffidence about the language tests may in fact have been partly justified given that, ideally speaking, many additional sources of evidence should be used to gather information about achievements in language learning (see Kohler, this issue). Intercultural understandings as we noted in the literature review, are not measured by our tests. Nor could we capture curriculum content that was specific to the learning program or the kinds of learning that might be achieved collaboratively with scaffolding from the teacher or other learners. The performance we elicited related only to the particular tasks on the tests we had designed. These were constrained in scope by the practicalities of having to administer them during class time without unduly intruding on day-to-day teaching activities.

The design of the tests was also influenced by the need to capture the full span of achievement from the beginning second language learner on the one hand through to the

fully-fledged native like learner with years of prior instruction through the medium of the target language on the other. For languages like Chinese, this extreme variation is the reality across Australian language classrooms (see Scarino, this issue and Scrimgeour, this issue). A single test for all students at each level was however needed for comparison purposes (i.e. to demonstrate language background and other time-related differences within the group). Thus, in spite of our best efforts to include tasks and items at different levels of difficulty there were sometimes floor and ceiling effects, such that less proficient learners were able to display very limited evidence of their learning, and more proficient learners may not have been able to show the full extent of their achievements.

The use of a common test for measuring learners also carries an assumption of unidimensionality (i.e. the notion that all learners can be lined up on a single scale). While this unidimensionality may be warranted in psychometric terms, it risks masking the fact that learners with different language and instructional backgrounds may differ not only in their level of proficiency but also in the nature and scope of their performance. While this variability was partially captured by comparing the performance of different learner subgroups across tasks or assessment criteria (e.g. see Kohler & Kim, this issue), giving the same assessment tasks to learners with vastly different learning and language background profiles can never do full justice to their different approaches to learning or to the different contexts and conditions through which language learning may have taken place. This is the central paradox our project, which required common measures to demonstrate the diversity present within the learner sample, and in so doing may have masked some of the critical differences between learners that it set out to reveal.

The involvement of experienced teachers in the process of drawing up descriptions of course helped to offset these limitations. These teachers were able to amplify the evidence from the performance samples with insights gleaned from their many years of working with similar learners. While always staying close to the evidence gathered, the teachers could extrapolate from it and explicate its meaning in a way that allowed for richer and more broad ranging descriptions than would have been possible based on the rather limited set of test or examination tasks used for the current study. Nevertheless, the descriptions we have developed should be seen as a starting point only and, in future projects, need to be supplemented with additional sources of empirical evidence of language learning achievement from different kinds of tasks, including those carried out collaboratively, via different media, including the internet, and from a broader range of schools and programs.

REFERENCES

Abbott, M. (2006). ESL reading strategies: Differences in Arabic and Mandarin speaker test performance. *Language Learning*, 56,633–670.

- Blondin, C., Candelier, M., Edelenbos, P., Johnstone, R., Kubanek-German, A. & Taeschner, T. (1998). Foreign languages in primary and pre-school education: context and outcomes. London: CILT.
- Bolger, P.A. & Zapata, G.C. (2011). Psycholinguistic approaches to language processing in heritage language speakers. *Heritage Language Journal*, 8(1), 1-29.
- Brown, A., Hill, K. & Iwashita, N. (2000). Is learner progress in LOTE learning comparable across languages? *Australian Review of Applied Linguistics*, 23(2), 35–60.
- Burstall, C. (1975). French in the primary school: The British experiment. *Canadian Modern Language Review*. Retrieved 22 November, 2009, from http://eric.ed.gov:80/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp.
- Carreira, M. (2004). Seeking explanatory adequacy: A dual approach to understanding the term "heritage language learner." *Heritage Language Journal*, 2(1), 1-25.
- Carreira, M. & Potowski, K. (2011). Pedagogical implications of experimental SNS research. Heritage Language Journal, 8(1), 134-147.
- Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, *1*, 131-151.
- Cleveland, H., Mangone, G. J., & Adams, J. C. (1960). *The overseas Americans: A report on Americans abroad*. New York: McGraw-Hill.
- Clyne, M. (1986). An early start: second language at primary school. Melbourne: River Seine.
- Clyne, M., Fernandez, S., Chen, I. & Summo-O'Connell, R. (1997). *Background speakers: Diversity and its management in LOTE programs*. Belconnen, ACT: The National Languages and Literacy Institute of Australia.
- Clyne, M., Jenkins, C., Chen, I. Y., Tsolalidou R., & Wallner, T. (1995). *Developing second language from primary school*. Canberra: NLLIA.
- Collins, L., Halter, R. H., Lightbown, P. M. & Spada, N. (1999) Time and the distribution of time in L2 instruction. *TESOL Quarterly*, *33*(4), 655-680.
- Corder S. P. (1981). Language distance and the magnitude of the language learning task. In Error analysis and interlanguage (pp. 95–102). Oxford: Oxford University Press.
- de Courcy, M. C. (2002). Learners' experiences of immersion education: Case studies of French and Chinese. Clevedon, UK: Multilingual Matters.
- DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. F. Freed, (Ed.), Foreign Language Acquisition Research and the Classroom (pp. 104-118). Lexington, MA: D.C. Heath.
- Dufon, M. & Churchill, E. (Eds). (2006). Language learners in study abroad contexts. London: Multilingual Matters.
- Elder, C. (1989). Drowning or waving? An evaluation of an Italian partial immersion program at a Victorian primary school. *Melbourne Papers in Applied Linguistics*, 1(2), 9-17.
- Elder, C. (1996). The effect of language background on 'foreign' language test performance: The case of Chinese, Italian and Modern Greek. *Language Learning*, 46(2), 233–82.
- Elder, C. (1997). The background speaker as learner of Italian, Modern Greek and Chinese: Implications for foreign language assessment. Doctoral dissertation, University of Melbourne.

- Elder, C. (2000a). Learner diversity and its implications for outcomes-based assessment. In C. Elder (Ed.). Defining standards and monitoring progress in languages other than English. Guest edited issue of the *Australian Review of Applied Linguistics*, 23(2), 36–61.
- Elder, C. (2000b). Outing the 'native speaker': The problem of diverse learner backgrounds in foreign language classrooms. *Language, Curriculum and Culture*, *13*(1), 86–108.
- Farmer, K. (2006). The Japanese bilingual program at Huntingdale Primary School, Melbourne. Babel, 41(2), 32–4, 38.
- Freed, B. F. (Ed.), (1995). Second language acquisition in a study abroad context. Amsterdam: Benjamins.
- Hill, K. (2012). Classroom-based assessment in the school foreign language classroom. Frankfurt am Main, Peter Lang.
- Hill, K., Davies, A., Oldfield, J. & Watson, N. (1997). Questioning an early start: the transition from primary to secondary foreign language learning. *Melbourne Papers in Language Testing*, 6(2): 21–36
- Iwashita, N. (this issue). Cross-linguistic influence as a factor in the written and oral production of school age learners of Japanese in Australia.
- Jarvis, S. & Pavlenko, E. (2008). Crosslinguistic influence in language and cognition. New York/London: Routledge.
- Johnstone, R. (2000). Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience. *Language Testing*, 17(2), 123–142.
- Johnstone, R. (2006). Review of research on language teaching, learning and policy published in 2004 and 2005. *Language Teaching*, 9(4), 1–27.
- Kagan, O. (2005). In support of a proficiency-based definition of heritage language learners: The case of Russian. The International Journal of Bilingual Education and Bilingualism, 8(2& 3), 213-221.
- Kim, H.-S. (2005). Processing strategies and transfer by heritage and non-heritage learners of Korean. Unpublished PhD dissertation, University of Hawaii at Manoa.
- Kim, J.-H., Montrul, S. & Yoon, J. (2009). Binding interpretations of Korean heritage speakers. Language Acquisition, 16, 3-55.
- Kim, S. H. O. (this issue). Learner background and the acquisition of discourse features of Korean in the Australian secondary school context.
- King, K. A. & Ennser-Kananen, J. (2012). Heritage languages and language policy. In C. Chapelle, (Ed.). The Encyclopedia of Applied Linguistics. Wiley-Blackwell. DOI: 10.1002/978140519843.
- Kohler (this issue). How does *time-on-task* affect the achievement of early and late starters of Indonesian in schools?
- Kondo-Brown, K. (2008). Issues and future agendas for teaching Chinese, Japanese and Korean heritage students. In K. Kondo-Brown & J. D. Brown, (Eds.), Teaching Chinese, Japanese and Korean heritage language students: Curriculum needs, materials, and assessment (pp. 17-43). New York: Lawrence Erbaum.

- Lado, R. (1957). Linguistics across cultures: applied linguistics for language teachers. Michigan, University of Michigan Press.
- Lapkin, S., Hart, D. & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. F. Freed, (Ed.), Second language acquisition in a study abroad context (pp. 67-94). Amsterdam: Benjamins.
- Llosa, L. (in press) Assessing heritage language learners. In A. Kunnan, (Ed.) *Companion to Language Assessment*. London: Wiley-Blackwell.
- Lorch, S., McNamara, T. & Eisokovitz, E. (1992). Late Hebrew immersion at Mt Scopus College, Melbourne: towards complete Hebrew fluency for Jewish day school students. *Language and Language Education*, 2(1), 1-29.
- Marriott, H. (1993). Acquiring sociolinguistic competence: Australian secondary students in Japan. *Journal of Asian Pacific Communication*, 4(4), 167–92.
- Molyneux, P. (2004). Pride and empowerment: Bilingually educated students reflect on their learning. Australian Language and Literacy Matters, 1(2), 4–10.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. Bilingualism: Language and Cognition, 5(1), 39-68.
- Montrul, S. & Perpiñán. S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. Heritage Language Journal, 8(1), 90-113.
- Odlin, T. (1989). Language Transfer: Cross-linguistic Influence in Language Learning. Cambridge: Cambridge University Press.
- O'Grady, W., Kwak, H-Y., Lee, O.-S., & Lee, M. (2011). An emergentist perspective on heritage language acquisition. *Studies in Second Language Acquisition*, *33*, 223-45.
- O'Grady, Lee, O-S & Lee, J-H, (2011). Practical and theoretical issues in the study of heritage language acquisition. *Heritage Language Journal*, 8(3), 23-39.
- Oostdam, R. & Van Toorenburg, H. (2002). 'Leuk is not enough': Het vraagstuk van de positionering van Engels inhebasisondervijs en de aansluitingmet het voortgezetonderwijs. Levende Talentijdschrift, 3(4), 3–17.
- Ozolins, U. (1993). The politics of language in Australia. Cambridge: Cambridge University Press.
- Polinsky, M. (2008). Without aspect. In G. Corbett & M. Noonan, (Eds.), *Case and grammatical relations* (pp. 263-82). Oxford: Oxford University Press.
- Polinsky, M. & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass*, 1(5), 368-395.
- Pufahl, I., Rhodes, N. C. & Christian, D. (2000). Foreign Language Teaching: What the United States Can Learn from other Countries. Eric Document ED-00-PO-4609.
- Rao, Z. (2001). Matching teaching styles with learning styles in East Asian contexts. *The Internet TESL Journal*, *VII*(7). Retrieved from http://iteslj.org/Techniques/Zhenhui-TeachingStyles.html.
- Rudd, K. (1994). Asian Languages and Australia's Economic Future: A report prepared for the Council of Australian Governments on a proposed national Asian languages/studies strategy for Australian schools. Brisbane: Queensland Government Printer.

- Scarino, A., Elder, C., Iwashita, N., Kim, S. H. O., Kohler, M. & Scrimgeour, A. (2011). *Student Achievement in Asian Languages Education. Full report.* Department of Education, Employment and Workplace Relations: Canberra.
- Scarino, A. (this issue). A rationale for acknowledging the diversity of learner achievements in learning particular languages in school education in Australia.
- Serrano, R. Tragant, E. & Llanes, A. (2012). A longitudinal analysis of the effects of one year abroad. The Canadian Modern Language Review, 68(2), 138-163.
- Slaughter, Y. (2007a). The study of Asian languages in two Australian states: Considerations for language-in-education policy and planning. Unpublished PhD thesis. Department of Linguistics and Applied Linguistics, University of Melbourne.
- Slaughter, Y. (2007b). The rise and fall of Indonesian in Australian schools: Implications for language policy and planning. *Asian Studies Review*, *31*, 301–322.
- Spada, N., & Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal*, 7(1), 11-32.
- Stern, H. H. (1985). The time factor and compact course development. *TESL Canada Journal*, 3(1), 13-27.
- Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), Heritage languages in America: Preserving a national resource (pp. 37–80). Washington, DC & McHenry, IL: Center for Applied Linguistics & Delta Systems.
- Victoria, Department of Education and Early Childhood Development (DEECD) (2010). Languages

 Other Than English in Victorian Government Schools 2009. Melbourne: DEECD. Retrieved 9

 September, 2012 from

 http://www.eduweb.vic.gov.au/edulibrary/public/LOTE/lotereport2009.pdf
- Vinjé, M. P. (1993). Balans van het Engels aanheteinde van de basisschool. Arnhem: CITO.
- Wiley, T. G. (2001). On defining heritage languages and their speakers. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 29–36). Washington, DC & McHenry, IL: Center for Applied Linguistics & Delta Systems.
- Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals*, *31*, 23–39.