

Understanding trajectories of academic literacy: How could this improve diagnostic assessment?

Bronwen Dyson

Faculty of Arts, Room 352 John Woolley Building, University of Sydney, Sydney NSW 2006, Australia

Email: bronwen.dyson@sydney.edu.au

(Received 13 August 2009; Published online 22 December 2009)

Given the rising interest in the English language development of international students in Australian universities, this paper considers the value of a developmental approach to the assessment of academic literacy. It outlines one of the criteria, Criterion D “Grammatical Correctness”, of the University of Sydney’s MASUS (Measuring the Academic Skills of University Students) Procedure (Bonanno & Jones, 2007) and discusses the need to underscore the validity of its assessment with developmental evidence. It then sketches a framework which has been used to measure language development, Processability Theory (Pienemann, 1998, 2005), and explores the applicability of this theory to the assessment of university students’ written English. By mapping the relationship between Criterion D and the oral development of two adolescent Chinese speaking students learning English as an Additional Language, the study reinforces the validity of scores on the criterion and its sub-criteria, the use of A (Appropriate) or NA (Not Appropriate) as measurement categories, and the overall score for grammatical performance. However, the findings suggest that the criterion’s “washback” to teaching could be fine-tuned by making the MASUS Procedure more “learner-sensitive”. The paper then discusses the study’s implications and limitations, focusing on the value and shortcomings of a developmental approach to academic literacy, particularly one concerned with grammatical development. The paper concludes that, despite the different foci of the empirical evidence and the MASUS Procedure, the findings suggest that an understanding of learner development could bolster two key features of language tests, namely validity and washback.

Key Words: academic literacy, diagnostic assessment, MASUS Procedure, Processability Theory, developmental trajectories

1. Introduction

Over the last decade, a range of procedures have been designed to diagnose the writing needs of international and local tertiary students (these are generically referred to as post-entry language assessments or PELAs). However, with growing interest in the English language development of international students (e.g. Bretag, 2007; Dunworth, 2009), these procedures are increasing in number and occupying a more significant role in policy making. Indeed, one of the ten “Good Practice Principles for English language proficiency for International students in Australian universities” proposed by the Australian Universities Quality Agency (AUQA) (2009) focuses specifically on the need for improved language assessment:

... Students’ English language development needs are diagnosed early in their studies and addressed, with ongoing opportunities for self-assessment.
(p. 4)

For two main reasons, this paper examines the PELA created by the University of Sydney: the MASUS (Measuring the Academic Skills of University Students) Procedure (Bonanno & Jones, 2007). Firstly, it is recognized as one of the best available due to the high validity of its assessment (Moore, 2008). Secondly, despite this validity, MASUS lacks developmental validation via data on the sequenced acquisition of features¹. Yet, such evidence would help MASUS achieve its purpose of diagnosing the academic literacy of university students, including the performance of the English as an Additional Language (EAL) students focused on in this article. Developmental validation would assist a wide range of participants in the diagnostic process. It would help raters be more reliable by providing them with detailed developmental trajectories against which they could check EAL student usage, such as emerging past tense marking on verbs. In addition, it would foster in academic staff a better understanding of the strengths and weaknesses in student scripts. For instance, when marking papers lecturers could note whether students were using both irregular verbs (*went, said* etc.) and regular past tense inflection (*listened, borrowed* etc.) or are limited to the earlier acquired irregular forms. Teachers delivering language and literacy support to EAL students would also benefit from developmental validation. They could target needs more precisely than at present by taking student strengths, such as irregular past tense verb usage, as the point of departure and moving onto weaknesses, such as the Past-ed. Perhaps most importantly, students could be made more aware of their position within trajectories via a developmental focus in diagnosis and support.

In order to ascertain whether MASUS has developmental validity, this paper explores the relationship between one of MASUS' five criteria (Criterion D "Grammatical correctness") and the longitudinal EAL development of two Chinese students. The framework adopted for this examination is Processability Theory (Pienemann, 1998, 2005), a theory of second language development. The paper argues that an understanding of developmental trajectories could bolster the validity of the MASUS Procedure and its influence on teaching, an effect commonly known as "washback" (Wigglesworth & Elder, 1996) and that this result has implications for other PELAs.

2. The MASUS procedure and Criterion D: "Grammatical correctness"

According to Dunworth's (2009) survey of PELAs in Australian universities, approximately one-third of these institutions have developed procedures. When compared to the characteristics of these measures, the MASUS Procedure (see Appendix A) is both typical and atypical. Typically, the procedure targets all students, is paper-based and assesses writing/reading (students read some disciplinary material and then write a response such as an essay). It also shares the main two objectives expressed – "to identify those who need English language support" and "to maintain/improve English language levels" (Dunworth, 2009, p. 4). Atypically, it tests grammar as well as writing/reading and, perhaps unusually although this is not clear from the survey results, it is discipline-specific since it tests skill in the genres of particular subject areas, albeit within a common framework².

As Appendix A shows, Criterion D focuses on grammatical skill by addressing the question "Is the message communicated without the interference of grammatical errors?" (Bonanno & Jones, 2007, p. 27). Like the other three criteria ("Use of source material", "Structure and development of answer" and "Academic writing style"), this criterion contains sub-criteria which, for Criterion D, are as follows:

- Clause structure follows recognizable and appropriate patterns of English;
- Correct subject/verb agreement;

¹ Language development is generally studied longitudinally but is sometimes approached cross-sectionally by collecting data on a group of learners with different levels of acquisition at a single point of time (e.g. Pienemann & Mackey, 1993).

² Dunworth's (2009) survey results for the content of PELAs are as follows: 10 writing, 7 test reading, 2 listening, 0 speaking, 2 vocabulary, 3 grammar, 1 Australian culture, 1 inference and 1 spelling.

- Consistent and appropriate tense choice, correctly formed; and
- Correct singular plural noun – agreement.

According to the student's control of the pertinent features, their writing is allocated an A (Appropriate) or NA (Not Appropriate) for each of these sub-criteria. A rating from 4 (excellent) to 1 (poor) is then given for the criterion as a whole. If a student receives a rating of 1 or 2, they are deemed to be "at risk" and in need of additional support.

Scoring on this criterion, as for the other MASUS criteria, has been validated. Holder, Jones, Robinson, and Krass (1999) found that all criteria except "use of source material" were predictors of academic performance and progress in a degree program. In addition, from his review of existing PELAs, Moore (2008) concluded that MASUS has a higher validity than many of its competitors. No doubt, the major reason for this is that the MASUS criteria were developed and validated against analyses of student writing which, in relation to grammatical usage for instance, revealed the patterns of "error" reflected in the sub-criteria (J. Jones, personal communication, May 25, 2009).

Although this analysis provides an important empirical basis for the criterion, it is not a developmental one. Yet, language development is a construct which MASUS seems to implicitly assess. In fact, language proficiency tests are based on the same assumption, as Brindley (1998) noted and as the AUQA (2009) report's dual emphasis on the "proficiency" and "language development" of international students further illustrates. Since an assumed construct requires validation for a test to have high validity (Hughes, 2003), Brindley (1998) suggests that proficiency tests use evidence of how people actually acquire second languages and theoretical frameworks which explain this evidence. While MASUS is a diagnostic procedure not a proficiency test, the same argument could be said to apply.

Of course, there are differences between proficiency tests and diagnostic tests which would affect how the developmental validation of each is approached. A crucial difference is target audience. In the context of instruction in English, proficiency tests assess EAL while diagnostic tests may assess both EAL and EMT (English as a Mother Tongue) students. Hence, proficiency tests need validation only on second language data while diagnostic tests need validation on both data sets. Although similarities have been found between the grammatical trajectories of first and second language users of English (e.g. Dulay & Burt, 1973, 1974), the developmental paths of the two groups are also likely to differ, especially at the tertiary level. By this time, EMT students have acquired the spoken mode of their first language, but may be moving through trajectories in the written mode (perhaps relating to recognition of clauses and sentences) while EAL students may be experiencing difficulties with both. Although the developmental validation of MASUS on EMT data merits future investigation, this study is limited to EAL data, and specifically oral data, in order to draw attention to the steps which EAL are traversing and the difficulties with spoken English which appear to underlie EAL students' written performance.

3. Processability theory: Hypotheses and findings

Processability Theory (PT) (Pienemann, 1998, 2005; Dyson, 2008, 2009) is one paradigm which has focused on how learners develop the grammar of second languages. PT is relevant to MASUS because it makes precise predictions about the domains which feature in the MASUS sub-criteria and captures development via language profiling (Pienemann & Kessler, 2007), an approach which interfaces with both diagnostic and proficiency testing.

PT hypothesises that learners acquire second (and additional) languages in developmental stages which are obligatory and qualitatively different skill-based speech processing levels (Pienemann, 1998). They do so because acquisition follows the same order as the processing of all speech production: first the word is processed, then the noun phrase, followed by the verb phrase, the main clause and finally the subordinate clause. Each of these components corresponds to a hierarchy of processing procedures, with one providing the prerequisites for the next:

- Lemma – Word in the mental lexicon
- Category – No exchange of grammatical information
- Phrasal – Noun Phrase information exchange
- Phrasal – Verb Phrase information exchange
- Sentence – Inter-phrasal information exchange
- Subordinate clause – Main and subordinate clause information exchange.

For EAL as for a range of other second languages, PT makes predictions for staged morphological and syntactic acquisition (Pienemann, 2005). The hypothesized order for EAL morphology corresponds well with the procedures³:

- Lexical morphology with no information exchange; e.g. PL-s (*dishes*)
- Phrasal morphology with a Noun Phrase information exchange; e.g. PL agreement (*those things*)
- Phrasal morphology with a Verb Phrase information exchange; e.g. Tense agreement (*have gone*)
- Inter-phrasal morphology with an inter-phrasal information exchange between the Noun Phrase subject and the verb; e.g. 3SG-s (*that man invites her*).

However, since the predicted syntax does not correspond to the procedures, it is accounted for by a separate explanation, the Topic Hypothesis which states that each word order stage corresponds to a new step in topic placement (Pienemann, 2005).

Dyson (2009) extends the hypothesis of processing procedures to account for problems in applying the Topic Hypothesis to English and to unify the explanation of EAL syntactic and morphological development. This account of the first five stages, those applicable to the learners in the present study, can be summarized as follows:

- Lemma – Word in the mental lexicon
- Category – Categorization of words and phrases, crucially single verb heads, results in the “lexical” SVO clause. Morphology, which may develop after the syntax, involves no grammatical exchange⁴
- Phrasal I – Categorization of the auxiliary (followed by a lexical verb) is a functional development at the clause level which is accompanied by new “phrasal” initial clausal positions and the remapping of *wh*- words. Noun Phrase morphology, which may develop after the syntax, involves a grammatical exchange within the Noun Phrase
- Phrasal II – The main clause, which remains “phrasal”, diversifies structurally, and displays additional remapping. Verb Phrase morphology, which may follow the syntax, involves a grammatical exchange within the Verb Phrase
- Sentence – Further diversification in the main clause results in full main clause sentence structure and further remapping. There is an inter-phrasal information exchange within the clause.

Although this process of development is constrained, PT does not see it as monolithic since it also predicts that learners vary in how they approach development (Pienemann, 1998, 2005). In recent work (Dyson, 2008, 2009), it has been suggested that these inter-learner differences can be conceptualized in terms of “developmental styles” with some learners more oriented towards the acquisition of word order (“lexical/syntactic learners”) and others towards morphology (“grammatical learners”).

³ Here, PL = plural, and 3SG = third person singular –s (on present tense verbs)

⁴ In Lexical Functional Grammar (Bresnan, 2001), a variant of universal grammar adopted by PT, “lexical” categories are distinguished from “functional” ones. Lexical’ categories, such as lexical verbs, tend to be semantically rich while “functional” ones, such as the category assigned to auxiliaries (I or Inflection), tend to be semantically empty and grammatical in nature.

While PT's focus on grammatical universals in the emergence of speech may seem a far cry from the assessment of non-English speaking background university students' written English, there is reason to believe that it is pertinent. The order hypothesized in PT has been supported by cross-sectional research on the EAL development of both instructed adults (Pienemann & Johnston, 1987; Pienemann, 2005) and children (Pienemann & Mackey, 1993; Pienemann, 2005) as well as longitudinal studies of university students' acquisition of a variety of other second languages, such as Arabic (Mansouri, 2005), Chinese (Zhang, 2005) and Japanese (Kawaguchi, 2005). Developmental stages are not only shared by learners of different ages but also by learners of varied language backgrounds, such as Vietnamese and Polish learners of English (Pienemann & Johnston, 1987). In addition, although there are important differences between the emergence or "onset" approach used in this paradigm and the appropriateness metric more typical of tertiary assessment, there are also some correspondences. For instance, accuracy studies of speech (e.g. Dulay & Burt, 1973, 1974; Anderson, 1978) report some similar orders in EAL morpheme acquisition to those predicted by PT. Studies in the related area of second language assessment have also shown the correlation between developmental stages and proficiency tests (Pienemann & Mackey, 1993) and the validity and practicality of profiling English developmental stages (Pienemann and Kessler, 2007). Certainly, one investigation of accuracy in spoken and written questions in a PT framework discovered "a noticeable disparity in results between oral and written tests" (Mansouri & Duffy, 2005, p. 93). The examples given, however, suggest that accuracy was not restricted to question syntax, leaving open the possibility of a closer relationship between the development of the two modes.

Noting the potential light which an exploration of the links between Criterion D and PT may shed on the diagnosis of academic literacy, this study addressed the following research question: Do the trajectories of grammatical development predicted by PT underscore the validity of Criterion D and, if so, what implications does this have for washback?

4. Research methods

4.1. Data collection

In order to validate MASUS developmentally, the present study takes a retrospective look at data collected in a larger doctoral investigation (Dyson, 2004). The original study was set in an Intensive Language Centre (ILC) which provided instruction in English, for a period of up to one year, to secondary aged students who had recently arrived in Sydney, Australia. For one academic year, the study tracked the grammatical development of six adolescents aged twelve to fourteen, who were initially placed in the junior class for students with a low level of English proficiency. The participants in this study were three males and females who were paired by their language background. The pairs were respectively speakers of Mandarin (and the Shanghai dialect), Arabic, and Bosnian (the first language)/German (the second language). The current investigation focuses on the Chinese-speaking pair, Daniel and Philomena, who were respectively thirteen and twelve years old when data collection started. The choice of Daniel and Philomena stems largely from the current importance of their language background in the Australian education system. Following a steady growth in the numbers of Chinese-speaking students, they are now one of the most numerous, if not the most numerous, group of international students in Australian universities (Zhang & Mi, 2009). Daniel and Philomena's development also illustrates some of the particular challenges facing Chinese learners of English, for instance their tendency towards analytic rather than inflected forms. For this reason, the conclusions regarding MASUS are based on these two students only. However, the overall similarities between Daniel and Philomena's trajectories and the other four (Dyson, 2004) suggest that their development has implications for the assessment of EAL students in general.

Following PT methodology, the data are oral and spontaneous since such data best capture the emergence of implicit grammatical form. To collect this data, the participants were audio-taped in a small quiet room as they carried out communicative tasks. Six samples of English speech, each lasting from 45 minutes to one hour, were collected over the academic year. Sample 1 was collected in March, Sample 2 in July, Sample 3 in August, Sample 4 in September, Sample 5 in

October and Sample 6 in December. The tasks used in these sessions were as follows: interview questions and story guessing (sample 1), picture guessing and picture description (sample 2), picture differences (sample 3), story guessing and narration (sample 4), story guessing (sample 5) and story guessing and narration (sample 6). Each task used sets of pictures which included contexts for all the grammatical structures predicted for EAL in PT. The audio-tapes were transcribed and analysed. It perhaps needs to be added that the study did not explore the students' explicit grammatical knowledge of English or, in any systematic way, their grammatical instruction. I only observed informally that the ILC curriculum emphasised the teaching of the major secondary school subjects, such as English, Science, and Maths, and that the English curriculum was skill-based with a small component relating to grammatical structure.

4.2. Data analysis

Taking the data analysis and findings in Dyson (2004, 2008, 2009) as its base, the current study investigated the relationship between these earlier results and MASUS Area Criterion D.

In the initial analysis, the transcriptions were analysed to ascertain which grammatical structures had emerged and in what order. In PT, emergence is defined as the first productive, or non-formulaic, use of a grammatical structure (Pienemann, 1998). Productivity is determined by two main means: the contexts and variability of a structure. Using the context-based approach, all tokens of PT structures were checked against their grammatical contexts. For example, the number of times a learner placed 3SG-s on a verb was contrasted with the number of times they used a base lexical verb with a third person noun subject. To establish emergence, a structure needed to be used non-formulaically at least once in the presence of four or more contexts (Pienemann, 1998)⁵. This measure was applied to both the syntactic and morphological PT structures. Using the contrast-based approach, tokens of structures which were marginal to or not predicted in PT⁶ were examined for their lexical and structural variability. For instance, in any one sample a learner employed 3SG-s on two different verbs, for example *goes* and *comes* (lexical contrasts) and also used these verbs in another form, for example *go* and *come* (structural contrasts). Emergence on this metric was determined when a feature was produced in four lexically and structurally different ways (Mansouri, 2005; Dyson, 2008, 2009). This second method has been previously applied to morphology to “neutralise” the effect of unanalysed entries in the learner’s lexicon” (e.g. Pienemann, 1998, p. 144; Mansouri, 2005). However, in this study it was also applied to syntax because some structures do not have clear contexts. For instance, Wh- Fronting is a context for other forms (Do/Aux2nd) but does not itself have a specific context⁷. These figures were then presented in a distributional table in which structures are ordered implicationally, meaning that a rule at a later stage implies the acquisition of a rule at the previous one (Pienemann, 1998)⁸.

⁵ Non-application is defined as no productive tokens in four contexts. For morphology, the productive tokens may be greater than zero but do not exhibit lexical and/or structural/morphological variation.

⁶ Some of these structures (Past Irregular, V-ing) were included in previous versions of the EAL stages (e.g. Pienemann & Johnston, 1987) but not the more recent ones (e.g. Pienemann, 1998, 2005), others (Past -ed, Wh- Fronting) are marginal members of the EAL stages since they have a place in PT but are not included in the EAL summary table (Pienemann, 2005) and the remainder have not been previously included (base verbs and nouns, Aux+ing, will, Aux+en, has/have).

7. Wh-Fronting = a constituent is fronted before the subject, verb and complement (in cases in which there is one) e.g. *What he doing?*. Do2nd = the auxiliary ‘do’ is placed in second position in direct questions, e.g. *what does her father want um him doing?*.

8. While the distributional tables of the two learners differ, they both reveal an implicational order in that the emergence of one structure at a more advanced stage implies the emergence of (at least) one structure at the previous stage.

To observe correspondences between the results of the emergence analysis and MASUS, these trajectories were then compared to the sub-criteria of Criterion D and tabulated. To render the table on clause structure manageable, it was necessary to omit some structures which were acquired such as negation and adverbs. All structures were then described and exemplified.

5. The developmental validity of Criterion D “Grammatical correctness”

By investigating the links between Criterion D “Grammatical correctness” and the learners’ development, three main sources of validity were revealed. First, the criterion itself and its sub-criteria identified major domains of EAL development as well as the difficulties Daniel and Philomena experienced in negotiating them. Second, the choice of A (Appropriate) or NA (Not Appropriate) for the sub-criteria reflected variation in these learners’ developmental routes. Finally, the overall score for grammatical performance revealed how their development fitted together into a “big picture”.

The ability of Criterion D and its sub-criteria to identify development and its difficulties was apparent from the students’ struggles with the grammatical domains identified by MASUS: the syntax of clauses and the morphological sub-criteria of subject/verb agreement, tense choice and singular plural noun-agreement. As Tables 1 to 4 reveal, in all these areas, Daniel and Philomena passed through developmental stages. For example, as Table 3 reveals about the development of singular-plural noun agreement, Daniel and Philomena first communicated via the semantics of “base nouns” (the word level) such as *my classmate* (when there was more than one classmate), before they added the inflectional ending Plural –s to nouns such as *dishes* (the category level), a step followed by the agreement between the determiner and the noun, for example *those things* (the phrasal level).

Because processing procedures are built up slowly around particular words, second language learning is often a lengthy business. As can be seen from Table 1, although these Chinese-speaking students were young and being educated in an English-medium environment, it took them an entire academic year to develop their clause structure from Subject - Verb - Object? (stage 2) to a limited range of full Wh- questions (stage 5), an achievement which falls short of the final stage (see Dyson, 2008 for a fuller account of question development). To give the reader a concrete idea of their struggles and achievements with questions, Figure 1 provides a snapshot of their questions in the final sample.

<u>Daniel</u>		
202.	D	yeah. where. come. f from?. and. mm. do you like here?
203.	I	mm
204.	D	and. do you want stay here forever?
205.	I	mm. mm.
206.	D	and. why. do you come here?
207.	I	and what would she say to that?
208.	D	and. I come here for. to sell my. country's goods. goods from the country. and because. my my. my country. or my home. is not rich
<u>Philomena</u>		
218.	P	where. where does her her father going?
219.	I	I think her father is still in China. I think that the. that the Chinese man is very very powerful I think like he's the king or the emperor of China. and I think that he that he put him in jail
220.	P	does this girl love um. the man. this man?
221.	I	um I think now that she does like him. yeah.
222.	P	why?
223.	I	because I think that she doesn't like this man. she doesn't want to marry this man ...
232.	P	she marry to him?
233.	I	yeah

Figure 1. Examples of the students’ questions (sample 6). (“I” = the interviewer)

Table 1. Stages in the development of clause structure. (“S” = sample; “–” = not acquired)

Stage	Question types	Examples	Description	Acquired by D	Acquired by P
Subordinate clause	Cancel inversion	<i>I wonder where he is.</i>	Learners acquire statement word order in indirect questions.	–	–
Sentence	Aux2nd Do2nd	<i>Where can he go?</i> <i>what does her father want um him doing?</i> <i>She don't know what does her father want um him doing?</i>	Learners place the auxiliary (do or another type) in second position in direct questions also overgeneralising this to indirect questions.	Do2nd S6	Do2nd S6
Sentence/ Phrasal II	Yes–no inversion (Y/N Inv) Copula inversion (CopInv)	<i>Can you put them away?</i> <i>How old is baby?</i> <i>Is she at home?</i>	Learners form wh-questions and yes–no questions via <i>inversion</i> , or remapping, of wh-words and copulas, as well as subjects and copulas/auxiliaries	CopInv S6	Y/N Inv S5 CopInv S5
Phrasal I	Wh-fronting (WhFr) Do-fronting (DoFr)	<i>What he doing</i> <i>Do you have this one?</i>	Learners question by <i>fronting</i> a constituent before the subject, verb and complement. In this way, they ask yes–no questions with an initial <i>do</i> and wh-questions with an initial wh-word.	WhFr S5 DoFr S5	WhFr S4 DoFr S3
Category	Rising intonation SVO? (Subject Verb Object/Compl ement)	<i>She want the noodle?</i>	Rising intonations remains the major questioning resource but at this point it is built onto SVO clauses.	SVO? S1	SVO? S1
Lemma	Word/s (W?)	<i>Yes?</i>	Second-language learners ask questions by adding rising intonation to single words or formulas.	W? S1	W? S1

The second reason for considering the MASUS assessment valid is that the identification of performance on the sub-criteria as either A (Appropriate) or N/A (Not Appropriate) detected variation similar to that in the learners' development. Philomena was oriented towards the development of word order not inflection. As Tables 1 to 3 reveal, by the final sample she had acquired sentence-level syntax (Do2nd questions) but only limited morphology: singular-plural noun agreement and one tense (Past irregular). In contrast, Daniel was oriented towards the development of morphology, where possible in tandem with syntax. As a result, by sample 6, he could productively use both the syntax and morphology (3SG-s) of this stage as well as a wider range of inflection: subject-verb agreement (has/have), singular-plural noun agreement, and several past tenses (Past Irregular, Past -ed, Aux+en, the last being a learner's form of verb phrase agreement). Hence, if achievement of the ultimate stage in a given domain is taken as the measure of "Appropriateness", in their sixth sample, both would receive As for singular-plural noun agreement, only Daniel would receive one for subject-verb agreement, and neither would receive them for tense and clause structure. While readers may see these differences as evidence that Daniel had a higher level of development than Philomena, the fact that both had reached the sentence level suggests that the variation is also a matter of "developmental style" in which Philomena was more "lexical/syntactic" in orientation while Daniel was more "grammatical".

The last point notwithstanding, the allocation of "A" or "N/A" to either of these learners is tentative since, unlike in MASUS, their development is measured in terms of emergence. To clarify the difference between emergence and appropriateness, let us look at how the emergence criterion measures acquisition of the 3rd person singular -s (the subject-verb agreement marker on present tense lexical verbs). In the final sample, Daniel acquired this structure because he used it at least twice on different verbs (lexical variety) which were also used in the base form (structural variety). In contrast, Philomena did not acquire this structure because she only used it once, which was insufficient to show lexical or structural variety. Yet, in the same sample both Daniel (49) and Philomena (42) did not mark third person singular -s on many eligible verbs. In other words, even though both learners were not inflecting many verbs "appropriately" for subject-verb agreement, there was a qualitative difference in acquisition. I shall return to the implications of these differences between MASUS and the emergence approach in the section on washback.

One final finding affirms the validity of MASUS scores. Its allocation of a total score to Criterion D captured the overall picture of the learners' development at a given point. Returning to the learners' sixth sample as an example and taking into consideration the difference between emergence and appropriateness, MASUS would award Daniel a "2" and Philomena a "1" for their level of acquisition respectively in two domains and a single domain. Thus, according to the "at risk" criterion cited earlier (a mark of 1 or 2), both students would be considered to fall in this category. Since this conclusion matches the considerable need of both students, we could supplement the MASUS global measure of "at risk" with an acquisition one for Criterion D, namely acquisition in less than two grammatical domains.

While the identification of Daniel and Philomena's needs would encourage language and literacy support, it falls short of providing an explanation of why they are "at risk" and hence a clear framework to guide the support. I would like to suggest that the concepts of stages and developmental styles could fill this gap. Both students were "at risk" because of their staged development: they were only part of the way through the trajectories of certain grammatical domains. However, only Philomena fell into this category because of her lexical/syntactic style: she had substantial gaps in her development of tenses and subject-verb agreement. In the next section, I shall explore how MASUS could accommodate this developmental perspective.

Table 2. Stages in the development of tense and subject-verb agreement. (“S” = sample; “-” = not acquired.) Note: There are no predictions relating to the development of tense and subject-verb agreement at the subordinate clause stage.

Stage	Verbs and their types of features	Examples	Description	Acquired by D	Acquired by P
Sentence	3SG-s	<i>so that man invites her to his house</i>	Learners mark subject/verb agreement on lexical verbs.	3SG-s S6	3SG-s –
Sentence	Has/have	<i>table has five pizza</i>	Learners express subject/verb agreement on main verbs, starting with ‘have’.	Has/have S5	Has/have –
Phrasal II	Tense agreement	<i>have gone</i>	Learners start to form Auxiliary & participle combinations, sometimes via Past tense marking on the auxiliary and main verb.	–	–
	Aux+en	<i>that mm. (c)hief was gone</i>		Aux+en S4	Aux+en –
Category	Future “will”	<i>she will get the baby</i>	Learners form the Future with will+ verb.	Future “will” S5	Future “will” S5
	Past -ed	<i>the woman sitted there</i>	They add inflection to verbs, first continuous aspect	Past -ed S4	Past -ed –
	Past irregular (Past Ir)	<i>he said it right</i>	(-ing with or without the auxiliary “is”) and then tense in the order Past Irreg, and Past-ed	PastIr S2	PastIr S4
	Auxiliary (Aux)+- ing	<i>what she wearing?</i>		Aux+ing S1	Aux+ing S2
	Base verbs (V)	<i>he want the noodle?</i>	Learners first use verbs in the base (uninflected) form, and rely on the semantics and syntax of the verb.	V S1	V S1

Table 3. Stages in the development of singular-plural noun agreement. (“S” = sample; “-” = not acquired.) Note: There are no predictions relating to the development of singular-plural noun agreement at the sentence and subordinate clause stages.

Stage	Nouns & their types of features	Examples	Description	Acquired by D	Acquired by P
Phrasal I	PL agreement (PLagr)	<i>those things</i>	Learners mark the plural on both the determiners and the nouns.	PLagr S4	PLagr S5
Category	PL-s	<i>two dish dishes</i>	Learners mark the plural on single nouns, not the determiner e.g. articles. Numerals may help learners acquire the plural.	PL-s S3	PL-s S5
	Base nouns (N)	<i>my classmate</i>	Learners first use nouns in the base form and rely on the semantics and syntax of the verb.	N S1	N S1

6. Ways to improve the developmental washback from Criterion D

The developmental evidence presented above implies that the MASUS Procedure could be adapted to improve its washback, or perhaps more appropriately, “washforward”. Washback, understood as the influence of a test on teaching, has for some time been considered a crucial dimension of any language test and remains the preferred term in the language testing field. (Wigglesworth & Elder, 1996; Hughes, 2003). Recently, the term “washforward” has been introduced to emphasise future test effects. This forward-looking perspective is certainly pertinent to MASUS since this procedure aims to guide the teaching of lecturers and writing advisors, and the learning of university students. Whichever concept one prefers, this study uncovers the kind of precise developmental information which Brindley (1998) argues is a crucial dimension of a positive “wash” from a test. Of course, this is not to deny that MASUS already has a beneficial pedagogical effect due to its diagnostic and discipline-specific emphasis, rather that currently it is unable to locate the specific points which students have attained in their developmental trajectories.

To raise Criterion D’s washback, a number of amendments could be made to MASUS. Firstly, as illustrated in Figure 2, the current “context-sensitive” approach of the “Collaborative procedure” (Bonanno & Jones, 2007) could be complemented by a more “learner-sensitive” one which compares the grammatical resources required by the target context with learner developmental paths, and then assesses and fosters development towards the desired level. For example, to supplement the initial step of establishing the academic literacy context, lecturers and writing advisors could note the grammar used in texts and, from the trajectories, predict the difficulties which EAL students may experience with it. This may prompt research on language structures which pose considerable difficulty to EAL learners but are not presently included in Criterion D (and are marginal to studies of stages), such as articles and voice (active and passive). Secondly, Tables 1 to 3 could be included in a new appendix with the categories of A (Appropriate) or N/A (Not Appropriate) substituted by the emergence categories of “Acquired” and “Not Acquired”. Thirdly, extra resources could be allocated to the rating process. Raters

would need training in how to apply these tables to writing via modules on the ESL structures and the application of the emergence criterion (see Pienemann and Kessler (n.d.) for an example of such material). Also, additional time would be required for the assessment of emergence.

The steps in the collaborative procedure with selected objectives	Ways of improving washback
1. Establish the context of literacy skills To identify: <ul style="list-style-type: none"> • desirable outcomes in terms of course/graduate literacy skills 	Identify the grammar most important for a context e.g. Past Irregular, Past-ed and Auxiliary+Past Participle in Science Lab reports and predict difficulties; e.g. emergence of Past Irregular & Past-ed but not Auxiliary+Past Participle (see Doughty & Varela, 1998).
2. Set up the literacy assessment task To determine: <ul style="list-style-type: none"> • content/ literacy skills required by task model answers 	Help select a task which assesses all these forms of the past tense e.g. laboratory reports with a range of past contexts (see Doughty & Varela, 1998)
3. Prepare for rating To include: <ul style="list-style-type: none"> • familiarization of raters with task and assessment criteria • exploration of rating problems 	Help raters to understand the difference between lexical (Past Irregular, Past-ed) and phrasal (Auxiliary+Past Participle) forms
4. Administer Literacy Diagnostic Task	
5. Debrief raters To explore difficulties with rating: <ul style="list-style-type: none"> • difficulties in applying criteria 	Help raters to understand the significance of learner forms e.g. Aux+en (was gone) initially seems to be a matching of Past+Past rather a passive and is indicated on a developmental point.
6. Respond to student literacy profiles Possible responses could include: <ul style="list-style-type: none"> • Integration of explicit development of literacy skills into the curriculum • Monitoring of progress of students 	Provide a specific framework for the integration of relevant grammatical aspects e.g. Science departments could include the development of the above past tenses in the lab report curriculum and assessment of lab reports could include use of past tenses.
7. Establishing support strategies for students identified as needing remedial assistance	Learning Centre and/or faculty-based academic writing support staff could offer workshops for “at risk” students on writing lab reports and past tense use.

Figure 2. Ways of improving washback from the collaborative procedure for establishing literacy diagnostic assessment within the context of specific disciplines.

7. Discussion and conclusion

By putting Criterion D of the MASUS Procedure under a developmental microscope, we have seen two main ways in which an understanding of developmental trajectories could improve

diagnostic assessment. Firstly, such an understanding could bolster the validity of a test: the analysis of the relationship between Criterion D and the grammatical development of two Chinese language background EAL students underscored the validity of the criterion and its sub-criteria and the ways in which performance on them is assessed. Secondly, a cognizance of developmental trajectories could heighten washback: an exploration of how this quality could be improved in MASUS revealed that information on trajectories would enable students' developmental points to be located and attended to via discipline-based assessment and teaching. These findings have implications for PELA in other universities.

One implication that can be drawn is that a developmental perspective has a legitimate place in literacy assessment. This conclusion echoes other current research. It is assumed in the AUQA (2009) Steering Committee's overriding goal – “the development of English language proficiency throughout students' studies” (p. 2), with the term “development” used synonymously with “language proficiency”. Although the terminology differs, it is also implied in Dunworth's (2009) summary of the second most commonly cited objective of Australian universities in introducing a PELA (reiterated from above): “to maintain/improve English language levels” (p. 5). Finally, the value of studying developmental trajectories in writing is apparent in work on secondary literacy, such as Christie and Derewianka (2009).

Admittedly, this other research has an orientation which differs from the one employed in PT: it is functional not structural. An emphasis on language function is evident in the AUQA (2009) report's definition of proficiency: “the ability of students to use the English language to make and communicate meaning in spoken and written contexts while completing their studies” (p. 1). A functional approach of a different nature, one based on Systemic Functional Linguistics (SFL), is adopted by Christie and Derewianka (2009) who define their framework as one which “illuminates how meaning is realized in language”(p. 1).

While a functional approach to development is more widespread than a structural one in the Australian context, this study adds to other evidence that a structural approach has much to offer assessment, in part because it can complement a functional one. Referring to the high correlation between proficiency tests and developmental stages for the same informants found in Pienemann and Mackey (1993), Pienemann and Kessler (2007) ask: “Does this mean that proficiency rating relies mainly on morphosyntax?” (p. 263). To investigate whether this is the case and how to bridge the “conceptual” gap between the narrow scope of developmental stages and the global scope of language proficiency tests, Pienemann and Kessler (2007) propose SLA studies of proficiency descriptors. While MASUS takes function not structure as its departure point, Criterion D also recognizes the role of structure and how it can be connected to the role of meaning in linguistic assessment. The question addressed by Criterion D exemplifies this point: “Is the message communicated without the interference of grammatical errors?”

This investigation has also focused attention on how the validity of assessment can be raised by the inclusion of a component concerned with the development of grammatical form. The strong relationship between Criterion D and the learners' paths suggests that assessment according to this criterion will have considerable construct validity, and hence test exactly what it aims to test (Wainer & Braun, 1988). In being able to arrive at this conclusion in a sense before the test event, we can also surmise that MASUS scores possess the two types of *a priori* validity: context-based and theory-based validity (Weir, 2005). The source of the former is the error analysis of student scripts by the MASUS creators and the latter could be seen as PT. Certainly, PT's cognitive framework differs theoretically from the socially oriented theory which inspires MASUS, SFL (Bonanno & Jones, 2007). Nevertheless, in the present socio-cognitive model of language testing (Weir, 2005), these theories could be seen as lending validity to different areas of MASUS assessment, PT to Criterion D and SFL to the other criteria.

Thus, PT could underscore the validity and especially construct validity essential to PELAs such as MASUS. Although tests must now possess a wider range of qualities than validity (Wigglesworth & Elder, 1998), a PELA must reveal the precise strengths and weaknesses of student work so that “at risk” students can be identified and appropriate support provided. MASUS can do this in the grammatical area because, as the SLA evidence underlines, the

selection of the criterion itself and its sub-criteria reflect major student difficulties. In one respect, however, the findings suggest that the validity of MASUS could be strengthened, thereby raising its diagnostic precision. Since the question which guides Criterion D (given immediately above) emphasizes the communication of meaning and limits structure to the role of interference via errors, raters may assess learner forms as “appropriate” if the meaning is sufficiently clear. Hence, areas which students need to develop may not be detected. To correct this imbalance, the question could be rephrased as follows: Is meaning communicated by the use of appropriate grammatical structures?

To increase their own validity, other PELAs could apply the developmental insights revealed in this research. Given its developmental validity, one option, pertinent to PELAs with no grammatical component would be to adopt Criterion D. Another would be to adapt it by reframing the question as suggested above. Yet another option would be to use data on trajectories to flesh out existing criteria, such as the one for Grammar and Vocabulary in DELNA (Diagnostic English Language Needs Assessment) (University of Auckland, 2007). This very general criterion translates to level descriptors which are fairly uninformative in relation to grammar, such as the following for Band 4 (Very limited user): “... Few sentence patterns are used correctly ...” (University of Auckland, 2007, p. 17).

A final conclusion that can be reached is that washback will increase via a fuller understanding of language development, including grammatical acquisition. By applying the learner data to the MASUS Procedure, this paper has shown that “learnable form” (Dyson, 2002), structures which students are ready to learn due to their developmental stage, can be integrated into the discipline-based, diagnostic and learning support process. This approach is based on the view that assessment – and resultant teaching – should emphasize “the students’ implicit knowledge alongside other components of communicative language ability” (Purpura, 2004, p. 253). Consequently, it is consistent with the current emphasis on discipline specific reading/writing (Dunworth, 2009). However, it also extends meaning-based grammatical assessment (Purpura, 2004) by highlighting the role of structure in the learning path (see Dyson, 2008 for an account of how meaning is also created in developmental stages).

This synthetic treatment of development would make washback more learner-sensitive than in two alternative approaches to grammar teaching in the tertiary context. In distinguishing its functional approach to proficiency (cited above) from one with “a narrow focus on grammar as a formal system concerned only with correct use of grammar and sentence structure”, the AUQA (2009, p. 1) report characterizes instruction in the latter as “remedial”⁹. In a similarly limited way, the assessment and teaching of “local grammatical forms” is seen as an “Editing” issue in the Diagnostic Assessment Profile produced by Allison, Cooley, Lewkowicz, and Nunan (1998, p. 204) to assess Hong Kong university students. Both approaches focus on “error”, neither one recognizing the use of erroneous alongside target and target-like forms as students grapple to develop the additional language grammar in communicative situations. Since MASUS also focuses on error, the suggested amendments to the MASUS “Collaborative Procedure” are designed to show how the pedagogy of grammatical structure can move away from errors and towards a more comprehensive assessment of students’ position within grammatical trajectories.

This study has limitations which mean that its observations should be followed up by more focused research. Since speaking and writing differ in their time constraints and hence the amount to which learners can “monitor” their grammatical usage, one major issue is whether the same trajectories are found in these two modes. Because emergence and appropriateness are different metrics of development, another issue is whether they are related. More generally, the question remains as to whether PELA should target the generic communicative ability of all students or purely the language development of EAL students (Dunworth, 2009). If the latter path is chosen, an additional concern is whether developmental assessment remains too narrow

⁹ The AUQA (2009) report contrasts “remedial” support with “developmental” support; i.e. actions to further develop students’ language proficiency.

due to its concern with morphology and syntax (Hudson, 1993). Despite these limitations, the study underscores the value of the MASUS Procedure, including its wise inclusion of grammar as a part of academic writing, and developmental studies.

Acknowledgements

I would like to acknowledge the valuable comments of Janet Jones, Sally Humphrey, Michele Zappavigna and two anonymous reviewers on drafts of this paper.

Appendix A. MASUS ASSESSMENT CRITERIA¹⁰

RATING SHEET (EXPERT LITERACY RATERS)

Cohort _____ Name _____ S.I.D. _____

KEY TO RATING:

- 4 = excellent / no problems / accurate / very appropriate A = appropriate
 3 = good / minor problems / mainly accurate / largely appropriate NA = not appropriate
 2 = only fair / some problems / often inaccurate / often inappropriate
 1 = poor / major problems / inaccurate / inappropriate

CRITERIA			
A. <i>Use of source material</i> - is information retrieval and processing of visual, verbal and numerical data correct and appropriate for the task?	4	3	2 1
<ul style="list-style-type: none"> ▪ most relevant data is employed ▪ use of irrelevant data is avoided ▪ visual and numerical data is interpreted correctly ▪ visual and numerical data is transferred correctly ▪ data is integrated with text ▪ text is free from plagiarism 	A		NA
B. <i>Structure and development of answer</i> - is the structure and development of the answer clear and generically appropriate to the question and its context?	4	3	2 1
<ul style="list-style-type: none"> ▪ genre is appropriate to the task ▪ clear focussed thesis statement ▪ choice of Theme and New reflects structure ▪ critical evaluation of evidence ▪ use of evidence consistent with thesis ▪ statement of conclusion which follows from argument / evaluation and relates to the thesis 	A		NA
C. <i>Academic writing style</i> - does the grammar conform to the patterns of written academic English appropriate for the task?	4	3	2 1
<ul style="list-style-type: none"> ▪ appropriate use of grammatical metaphor and nominal group structure ▪ appropriate use of interpersonal metaphor ▪ demonstrated control of appropriate modality 	A		NA

¹⁰ Bonanno & Jones (2007, p. 14).

<ul style="list-style-type: none"> ▪ demonstrated control of cohesive devices - reference chains, textual reference ▪ demonstrated control of taxonomic relations ▪ appropriate choice of lexis 		
D. Grammatical correctness - do grammatical errors interfere with communicating the message?	4	3 2 1
<ul style="list-style-type: none"> ▪ clause structure follows recognisable and appropriate patterns of English ▪ correct subject/verb agreement ▪ consistent and appropriate tense choice, correctly formed ▪ correct singular / plural noun agreement 	A	NA
E. Qualities of presentation	not rated	
<ul style="list-style-type: none"> ▪ spelling generally correct ▪ handwriting legible ▪ paragraphing reflects essay structure 	A	NA

References

- Australian Universities Quality Agency. (2009). *Good Practice Principles for English language proficiency for International students in Australian universities – Final Report Part 1*. Australian Universities Quality Agency.
- Allison, D., Cooley, L., Lewkowicz, J., & Nunan, D. (1998). Dissertation writing in action. *English for Specific Purposes*, 17, 199-217.
- Anderson, R. (1978). An implicational model for second language research. *Language Learning*, 28, 221-282.
- Bonanno, H., & Jones, J. (2007). *The MASUS Procedure: measuring the academic skills of university students*. Sydney: Learning Centre, The University of Sydney.
- Bretag, T. (2007). The Emperor's new clothes: Yes, there is a link between English language competence and language standards. *People and Place*, 15(10), 13-21.
- Brindley, G. (1998). Describing language development? Rating scales and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between Second Language Acquisition and Language testing research* (pp. 112-140). Cambridge: Cambridge University Press.
- Christie, F., & Derewianka, B. (2009). *School discourse: Learning to write across the years of schooling*. London: Continuum.
- DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty and M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-348). Malden, MA: Blackwell Publishing.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty and J. Williams (Eds.), *Focus on Form in classroom Second Language Acquisition* (pp. 114-138). Cambridge: Cambridge University Press.
- Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, 23, 245-258.
- Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, 24, 37-53.
- Dunworth, K. (2009). An investigation into post-entry English language assessment in Australian universities. *Journal of Academic Language and Learning*, 3(1), A1-A13.

- Dyson, B. (2002). Focus on learnable form in a communicative context: A framework for second language acquisition in the classroom. *Australian Review of Applied Linguistics*, 25(1), 53–71.
- Dyson, B. (2004). *Developmental style in second language processing: A study of inter-learner variation in the acquisition of English as a Second Language*. Doctoral dissertation, University of Western Sydney.
- Dyson, B. (2008). What we learn from questions: ESL question development and its implications for language assessment, *Prospect*, 23(1), 16-27.
- Dyson, B. (2009). Processability Theory and the role of morphology in ESL development: A longitudinal study. *Second Language Research*, 25(3), 355-376.
- Holder, G. M., Jones, J., Robinson, R. A., & Krass, I. (1999). Academic literacy skills and progression rates amongst pharmacy students. *Higher Education Research and Development*, 18(1), 19-30.
- Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, 15, 461-494.
- Hughes, A., (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Kawaguchi, S. (2005). Argument structure and syntactic development in Japanese as a second language. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 253-298). Amsterdam: John Benjamins Publishing.
- Mansouri, F. (2005). Agreement morphology in Arabic as a second language. In M. Pienemann (Ed.). *Cross-linguistic aspects of Processability Theory* (pp. 117-253). Amsterdam: John Benjamins Publishing.
- Mansouri, F., & Duffy, L. (2005). The pedagogic effectiveness of developmental readiness in ESL grammar instruction. *Australian Review of Applied Linguistics*, 28(1), 81-99.
- Moore, T. (2008). *Diagnostic language and literacy assessment policy: Balancing system and lifeworld*. Presentation at NSW/ACT AALL Region meeting, University of Sydney, 25 September 2008.
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins Publishing.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins Publishing Company.
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying Second Language Acquisition Research* (pp. 45-141). Adelaide: National Curriculum Research Centre, Adult Migrant Education Program.
- Pienemann, M., & Kessler, J.-U. (2007). Measuring bilingualism. In P. Auer & L. Wei (Eds.). *Handbook of applied linguistics. Vol. 5: Multilingualism and multilingual communication* (pp. 247-275). Berlin/New York: Mouton/de Gruyter.
- Pienemann, M., & Kessler, J.-U. (n.d.) *Rapid Profile English flyer. Rapid Profile – linguistic profiling*. University of Paderborn, Germany English Department. Retrieved April 21, 2008, from <http://groups.uni-paderborn.de/rapidprofile>
- Pienemann, M., & Mackey, A. (1993). An empirical study of children's EAL development and Rapid Profile. In P. McKay (Ed.), *EAL development. Language and literacy in schools*. Vol. 2 (pp. 115-259). Commonwealth of Australia and National Languages and Literacy Institute of Australia.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.
- University of Auckland (in conjunction with the Language Testing Research Centre at the University of Melbourne). (2007). *Diagnostic English Language Needs Assessment*

Handbook for Candidates. Retrieved August 7, 2009, from <http://www.delna.auckland.ac.nz/about.php>.

- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Hampshire: Palgrave Macmillan.
- Wainer, H., & Braun, H. I. (Eds.). (1988). *Test validity*. Hillsdale, New Jersey: Lawrence E. Erlbaum Associates.
- Wigglesworth, G., & Elder, C. (1996). Perspectives on the language testing cycle: setting the scene. *Australian Review of Applied Linguistics*, S(13), 1-12.
- Zhang, Y. (2005). Processing and formal instruction in the L2 acquisition of five Chinese grammatical morphemes. In M. Pienemann (Ed.). *Cross-linguistic aspects of Processability Theory* (pp. 155-177). Amsterdam: John Benjamins Publishing.
- Zhang, Y., & Mi, Y. (2009). Another look at language difficulties of international students. *Journal of Studies in International Education*, 20(10), 1-18.