Zapata-Rivera, D., VanWinkle, W., Shute, V. J., Underwood, J. S., & Bauer, M. I. (2007). English ABLE. International Conference of Artificial Intelligence in Education. 8pp.

English ABLE

Diego ZAPATA-RIVERA, Waverely VANWINKLE, Valerie SHUTE, Jody S. UNDERWOOD, and Malcolm BAUER
Educational Testing Service, Princeton, NJ 08541 USA

Abstract. Assessment-Based Learning Environments (ABLE) make use of assessment information coming from a variety of sources (e.g., formative and summative) to guide instruction. We have developed English ABLE, an assessment based learning environment designed to help English language learners (ELLs) learn about English grammar. Main features of English ABLE include: Item/task reuse (900 enhanced TOEFL® items were recalibrated based on data from all native Spanish speakers who have taken the test), a Bayesian psychometric student model that makes use of item statistics, adaptive feedback, adaptive sequencing of tasks, pedagogical agents (Dr. Grammar, Jorge and Carmen), and an indirectly visible student model. This paper presents English ABLE and reports on some preliminary results from a study that involved 149 native Spanish-speaking ELLs.

1. Introduction

Researchers in the area of second language acquisition have explored the types of errors that language learners make in relation to their native language (L1), target language (L2), and proficiency level [1-5]. Cowan [1], for example, reports on four major causes for errors made by second language learners: interference from native language, learning strategies that are similar to first language, context of the learning situation, and emotional factors connected with pressure of communication. More recently, Schuster & Burckett-Picker [3] identify six strategies for learning a second language in terms of interlanguage (a dynamic variant of the target language that evolves from an existing knowledge base of the native language): direct use of native language, negative transfer, reduction of grammatical redundancy, simplification, positive transfer, and overgeneralization. Bull [4] also emphasizes the importance of understanding transfer errors and utilizing characteristics of the learner's native language when designing intelligent computer-assisted language learning (ICALL) systems. ICALL systems that make use of student models include: Mr. Collins [5], and I-PETER [6].

We have designed an Assessment-Based Learning Environment (English ABLE) that makes use of student test data gathered from native Spanish speakers and expert opinions to create a Bayesian student model that allows for the development of various advanced adaptive technologies. Specifically, English ABLE features adaptive feedback, adaptive sequencing of enhanced TOEFL® tasks, pedagogical agents, and an indirectly visible student model.

This paper is structured as follows: Section 1 presents English ABLE (i.e., task reuse, Bayesian student model, pedagogical agents and the indirectly visible student model). Section 2 reports on a preliminary analysis of data gathered in November, 2006. We conclude with some final remarks and our plans for future work in this area.

2. English ABLE

English ABLE is an Assessment-Based Learning Environment for English grammar. Assessment-based learning environments make use of assessment information to guide instruction. English ABLE demonstrates the reuse of existing high-stakes tasks in lower stakes learning contexts. English ABLE currently draws upon a database of TOEFL® CBT tasks to create new packages of enhanced tasks targeted towards particular component ELL skills. It presents these packages to learners within an adaptive, scaffolded learning environment to help students master aspects of English grammar. In English ABLE, students try to help a virtual student (Carmen or Jorge) learn English by correcting this student's writing from a notebook of facts (sentences – enhanced TOEFL® tasks). Supplemental educational materials about specific grammatical structures are offered by a virtual tutor (Dr. Grammar). Figure 1 shows a screenshot of English ABLE. The student is helping Jorge find grammatical errors on several sentences. However, he/she has not been successful at finding the errors (only one out of four sentences has been answered correctly). Jorge's knowledge levels, which are also the student's knowledge levels (indirectly visible student model, see section 1.4), show a lack of knowledge for agreement. Jorge seems confused and expresses it ("*I don't understand how to make the verb agree with the rest of the sentence.*"). Dr. Grammar offers immediate verification feedback ("*I see you have selected 'created'. However, this part of the sentence is correct.*"), and additional adaptive instructional feedback (i.e., rules, procedures, examples and definitions).

Figure 1. A screenshot of English ABLE

2.1. Task reuse

Existing test items (tasks) are made available to new applications through the use of metadata that is added manually or automatically for a particular purpose. Task statistics (e.g., difficulty and discrimination parameters), for example, can be used to inform on-line instruction. Additional pieces of metadata include links to relevant learning objects, and a map of concepts (e.g., proficiency map). English ABLE includes approximately 900 enhanced TOEFL® tasks. Packages of enhanced tasks can be used to support a variety of applications including learning environments.

2.2. Bayesian student model

English grammar can be divided into three main categories: use, form and meaning [7, 8, 9]. Working with experts we have elicited an initial Bayesian structure for a student model (see Figure 2). This Bayesian structure can be used to capture and propagate evidence of student knowledge regarding some aspects of English grammar. The current structure of the Bayesian student model deals with English grammar form (although it could be extended to cover use and meaning). Three sentence–level grammatical categories (i.e., wrong form, agreement, and omission and inclusion) have been chosen based upon a difficulty analysis that was performed using student data (i.e. student responses) from only native Spanish speakers. These three sentence-level grammatical categories are further divided into low-level sub-categories (leaf-nodes) according to parts of speech (e.g., agreement has been divided into 3 leaf nodes: noun agreement, verb agreement and pronoun agreement). Leaf-nodes are linked to 2 main knowledge areas (i.e., individual parts of speech: noun, verb and pronoun, and sentence-level grammatical categories) (see Figure 3). Preliminary difficulty analysis plus data from experts were used to generate prior and conditional probabilities for the latent structure. Experts used a qualitative inspired method to produce probability values based on estimates of the strength of the relationship between any two variables in the model [10]. Each task is attached to a single category using existing classification metadata and corresponding IRT (Item Response Theory) parameters (see below).

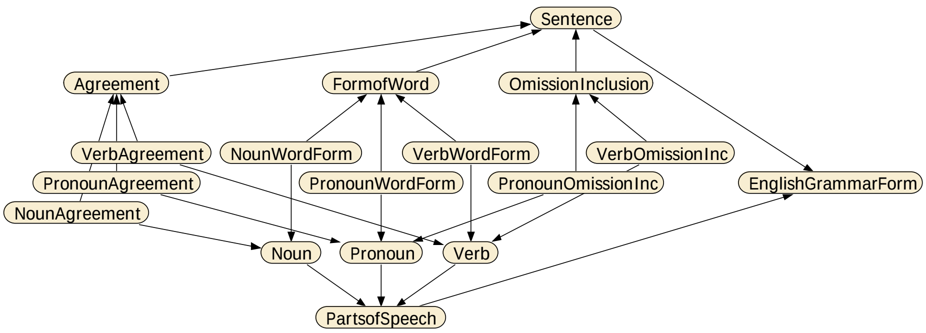

Figure 2. Bayesian student model structure

Tasks were recalibrated based on data from all native Spanish speakers who took the test. Each task is connected to a single leaf-node (proficiency) using the resulting IRT parameters. The IRT-2PL model is described by the following formula:

$$\Pr(task_i = correct \mid proficiency_j) = \frac{1}{1+(e^{-1.701*a(proficiency_j - b)})} ,$$

Where b is the difficulty parameter ($-3 <= b <= +3$, typical values for b), a is the discrimination parameter ($-2.80 <= a <= +2.80$, typical values for a), and $Proficiency_j$ represents an ability level (continuous proficiency variables were discretized using the following ability values: Advanced = 0.96, IntermediateAdvanced = 0, and

Intermediate = -0.96. These values come from quantiles of a normal distribution [11]). Table 1 shows the resulting conditional probability table of Task 2 (*a*=1.5, and *b*= 0.4).

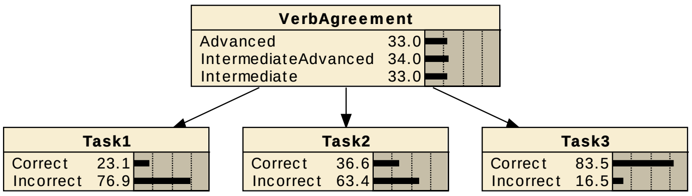

Figure 3. Various tasks attached to VerbAgreement using IRT parameters and organized by difficulty

Table 1: Conditional probability table for Task 2 (*a*=1.5, *b*= 0.4)

	Pr(Task2 | VerbAgreement)	
VerbAgreement	**Correct**	**Incorrect**
Advanced	0.807	0.193
IntermediateAdvanced	0.265	0.735
Intermediate	0.030	0.970

As the student makes progress (i.e., answers additional tasks), more tasks/nodes are dynamically added to the model. Observed values per task (i.e., correct or incorrect) provide evidence (as defined by its conditional probability table) to update the student model. Pedagogical agents query the Bayesian student model to provide adaptive feedback, adaptive sequencing of items, and adaptive behavior.

2.3. Pedagogical Agents

Pedagogical agents (e.g., [12, 13, 14]) have been used to facilitate learning by supporting human-like interaction with computer-based systems. Pedagogical agents can act as virtual peers or virtual tutors. Pedagogical agents can model human emotions and use this information to facilitate learning (e.g., [15, 16]). Teachable agents [16], an interesting variant of pedagogical agents, have been used to facilitate student learning. The student's role in these environments is to teach an artificial student how to act in the simulated environment.

Students in English ABLE are asked to help a pedagogical agent (i.e., Carmen and Jorge) find grammar errors. Carmen and Jorge "learn" based on the student's performance. Students can see how much the pedagogical agent knows about a particular concept by looking at the indirectly visible student model (i.e., knowledge levels, see Section 1.4), and by observing Carmen's and Jorge's changes in emotional states and associated utterances.

Carmen and Jorge are able to express basic emotions (see Figure 4), which are triggered by a list of predefined rules. These rules take into account recent performance (e.g., if three sentences attached to the same grammatical structure have been answered correctly in a row, then *emotion* = *happy*) and changes to the student model (e.g., if $\Pr(proficiency_j \mid evidence) \leq threshold\ value$, then *emotion* = *worried*). Utterances for pedagogical agents are dynamically generated and attached to particular emotions. Pedagogical agents in English ABLE are aimed at keeping students motivated.

Dr. Grammar (see Figure 1) provides adaptive instructional feedback (i.e., rules, procedures, examples and definitions) based on the student model. Feedback is not handcrafted at the item level but instead is keyed to grammatical categories, students' native language, and estimated skill level. Amount and type of information presented to the student varies according to the student's knowledge level (i.e., intermediate students receive more detailed information compared to intermediate-advanced and advanced students).

Happy	Confused	Worried	Neutral
			
			

Figure 4. Jorge's and Carmen's emotional states

The difficulty level of the next item (within each grammatical category) is chosen based on short-term past performance and current state of the student model (long-term) following a CAT-like (computer-adaptive testing) approach. Dr. Grammar changes grammatical categories when the marginal probability of $proficiency_i = state_j$ reaches a particular probability threshold value, e.g., 0.80). The next category is selected based on a predefined sequence of categories obtained through preliminary difficulty analysis and mutual information analysis using the Bayesian model.

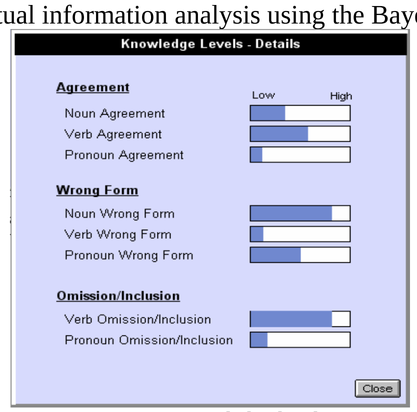

Figure 5. Knowledge levels

2.4. Indirectly visible student model

English ABLE supports indirect inspection of the student model (i.e., knowledge levels in Figures 1 and 5). We consider that exploring one's student model via a pedagogical agent is less intimidating and has the potential to foster student learning without the possible negative effects on self-esteem and motivation, especially for those students who are having a hard time with the system [18]. Previous research on inspecting Bayesian student models through the use of guiding artificial agents showed that agents can facilitate student interaction with the model by helping students navigate and find conflicting nodes. Guided agent interaction was linked to higher levels of student reflection [19].

The length of each bar is determined based on the probability distribution of each node/proficiency using the following formula:

$Length_i = \left[\sum_{j}^{n} C_j \Pr(proficiency_i = state_j) \right] / n$, where Cj is a constant numerical value assigned to each state of a node based on its proficiency level (i.e., Intermediate = 0, IntermediateAdvanced = 1, and Advanced = 2) and n is the index of the highest proficiency state (e.g., n = 2, in this case). This produces an Expected A Posteriori (EAP) score that ranges from 0 to 1.

3. The experiment

A study focused on examining usability issues and its effectiveness at helping native Spanish speakers learn English grammar was carried out in November, 2006. The current analysis includes data from 149 participants assigned to 3 conditions (see table 3). Participant institutions include 1 high school, 4 community colleges, 1 college/university, and 2 adult learning centers in four different states (i.e., NJ, PA, NY, and MD). The majority of our sample was comprised of community college students. Participants' ages ranged from 15 to 73 years old with a mean age of 33 and a mode of 21 years of age (all the participants were native Spanish speakers).

3.1. Experimental design

Each session consisted of a 20-minute pretest, a 1-hour intervention (*conditions 1*, *2* or *3*), a 20-minute posttest, and a 10-minute usability survey. Conditions 1, 2 and 3 had the features described in the table below included in the learning tool to different degrees: interface, student model, and pedagogical agents.

Table 2: Experimental design

Features	Condition 1 (control)	Condition 2	Condition 3
Interface	Test preparation environment (text-based)	English ABLE simple	English ABLE enhanced (visible student model, hints and making corrections)
Student Model	None	None	Bayesian student model
Pedagogical Agents	None Verification feedback available. Predefined, pedagogically-sound task sequencing.	Carmen/Jorge (limited interactions based on short-term performance – last 3 sentences) Dr. Grammar (verification feedback only, predefined, pedagogically-sound task sequencing)	Carmen/Jorge (interactions based on student model – short and long-term) Dr. Grammar (verification and adaptive instructional feedback and adaptive task sequencing based on the student model)

3.2. Results

This section presents some preliminary results.

Usability information –Pedagogical Agents

Participants (88% or more across all three conditions) thought that the feedback provided by the system was useful. Regarding Jorge's and Carmen's visible student model (*condition 3*), most participants (88%) stated that they understood the knowledge levels, 86% thought that the knowledge levels were useful, and 86% agreed that the knowledge levels helped them understand what Jorge/Carmen knew. In general, participants thought that they had learned a lot from using the software (81% or more across all three conditions).

Participants in *conditions 2* and *3* agreed with the following statements: (a) "*I liked helping Carmen/Jorge find grammar errors*" (95% and 89%, conditions 2 and 3 respectively), (b) "*Carmen's/Jorge's comments were useful*" (75% and 81%, respectively), (c) "*Helping Carmen/Jorge motivated me to keep going*" (88% and 92%, respectively), (d) "*I have helped Carmen/Jorge a lot by finding the grammar errors*" (75% and 71%, respectively), (e) "*I have learned by helping the Carmen/Jorge with his/her sentences*" (92% and 89%, respectively), (f) "*The feedback provided by Dr. Grammar helped me learn*" (96% and 78%, respectively), and (g) "*I think Carmen and Jorge liked my help*" (83% and 79%, respectively). Because of the exploratory nature of our usability study, we did not do statistical analysis on these data.

Learning effects

We report on the following three research questions:

1) *Condition.* What is the effect of increasing the number of features (*conditions 1*, *2* and *3*) in the English ABLE environment on learning outcome?

2) *ESL level*. What is the relationship between ESL level on learning outcome in English ABLE?

3) *Condition x ESL Level.* Are any of the conditions more or less supportive of learners at different levels of their language proficiency?

To answer these questions, we computed an ANCOVA using *posttest correct* as our dependent variable, and *condition* (*1*, *2*, and *3*) and *ESL level* (Intermediate and Advanced)[1] as our independent variables. To control for differences in (a) incoming knowledge, and (b) number of posttest items actually solved, we used *pretest correct* and *posttest items answered* respectively as covariates in the analysis. The results of the ANCOVA predicting learning outcome were as follows: (a) significant main effect of *condition* ($F_{2,\ 129} = 3.42$, $p < 0.05$), (b) significant main effect of *ESL level* ($F_{1,\ 129} = 12.18$, $p < 0.01$), and (c) no significant interaction between *condition* and *ESL level*. Estimated means, standard errors, and sample sizes for the posttest-correct data, separated by condition and ESL levels are presented in Table 3.

Table 3: Estimated Marginal Means for Posttest Correct (with Standard Error and N)

Condition	Intermediate	Advanced
1. Test preparation	**13.6** (0.7, 36)	**15.2** (1.3, 12)
2. English ABLE simple	**14.1** (0.8, 26)	**16.3** (0.9, 24)
3. English ABLE enhanced	**14.6** (0.8, 30)	**19.8** (1.4, 9)

Regarding the significant main effect of condition on outcome, Table 3 suggests that more sophisticated versions of the program yield progressively better learning outcomes. To test this assertion, we computed a planned comparison (Least Significant Difference) on these data and results showed that *condition 1* was significantly different from *condition 3* (mean difference = -2.8, *SE* = 1.1, $p < 0.05$), and *condition 2* was marginally different from *condition 3* (mean difference = -2.0, *SE* = 1.0, $p = 0.052$). And while *conditions 1* and *2* were not significantly different from each other, the trend is in the right direction. Finally, the main effect of ESL level indicates that the more advanced students demonstrated greater learning, overall, compared to their intermediate counterparts.

4. Conclusions and future work

Preliminary findings reported in this paper show that English ABLE has potential for facilitating student motivation and learning. We just started exploring the benefits of

[1] The sample size for our self-reported Beginners was very small ($n = 12$), particularly within some of the conditions (e.g., test prep $n = 3$ and English ABLE simple $n = 2$), so we eliminated them from the current analysis.

English ABLE. A longer study would provide more information regarding the adaptive, persistent features of English ABLE enhanced (*condition 3*). Additional reflection tasks can also be included, which would make the profile of pedagogical agents more prominent. Initial results demonstrate that English ABLE has a role to play in helping ELLs learn English grammar. Future work would include adapting the Bayesian model and feedback to various native languages.

References

[1] Cowan, J.R. (1976, July). Toward a psychological theory of interference in second language learning. Paper presented at the Conference on Second Language Learning and Teaching, New York, NY.

[2] Schuster, E. (1986). The role of native grammars in correcting errors in second language learning. *Computational Intelligence*, 2(2), 93-98.

[3] Schuster, E., Burckett-Picker, J. (1996). Interlanguage errors becoming the target language through student modeling. *Fifth International Conference on User Modeling*, Kailua-Kona, Hawaii.

[4] Bull, S. (1995). Handling native and non-native language transfer in CALL: Theory and practice. R. Wakely,

[5] Bull S. & Pain H. (1995), "'Did I say what I think I said, and do you agree with me?': Inspecting and Questioning the Student Model", *Proceedings of World Conference on Artificial Intelligence in Education (AACE)*, Charlottesville, VA, 501-508.

[6] Read, T., Barcena, E., Barros, B., & Verdejo, F. (2002). Adaptive modelling of student diagnosis and material selection for on-line language learning. [On-line version]. *Journal of Intelligent & Fuzzy Systems*, 12, 135-149.

[7] Celce-Murcia, M., & Larsen-Freeman, D. (1999). The Grammar Book (2nd ed.). Heinle & Heinle Publishers.

[8] Greenbraum, S. (1996). The Oxford English Grammar. New York: Oxford press.

[9] Gucker, P.(1966). Essential English Grammar. NJ: Dover, Publications. New York and London: Dover Publications.

[10] Daniel, B., Zapata-Rivera, J.D., McCalla, G. (2003) A Bayesian Computational Model of Social Capital in Virtual Communities. *Proceedings of the First International Conference on Communities and Technologies; C&T 2003*. Kluwer Academic Publishers. ISBN 1-4020-1611-5.

[11] Almond, R., Dibello, L., Jenkins, F., Mislevy, R., Senturk, D., Steinberg, L., and Yan, D. (2001) Models for conditional probability tables in educational assessment. *In T. Jaakkola and T. Richardson, editors, Artificial Intelligence and Statistics*, 137–143.

[12] Chan, T. W. & Baskin, A. B. (1990). Learning Companion Systems. In Frasson, C., and Gauthier, G. (eds.) *Intelligent Tutoring Systems: At the crossroads of AI and Education*. Ablex Pub., 6-33.

[13] Graesser, A. C., Person, N., Harter, D., & TRG (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 257-279.

[14] Johnson, W. L.; Rickel, J. W.; and Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, 11(1), 47-78.

[15] Picard, R.W. (1997). Affective computing. Cambridge, MA: MIT Press.

[16] Biswas, G., Schwartz, D., Bransford, J., and the Teachable Agent Group at Vanderbilt (TAG-V) (2001). Technology support for complex problem solving: From SAD environments to AI. *In K. D. Forbus and P. J. Feltovich (Eds.), Smart machines in education: The coming revolution in educational technology.* Menlo Park, CA: AAAI/MIT Press. 71-97.

[17] Nkambou, R., Laporte, Y., Yatchou, R., and Gouardères, G. (2003). Embodied emotional agent in intelligent training system. *In Recent Advances in intelligent Paradigms and Applications, J. Kacprzyk, A. Abraham, and L. C. Jain, Eds. Studies In Fuzziness And Soft Computing.* Physica-Verlag Heidelberg, 235-253.

[18] Bull, S., Mangat, M., Mabbott, A., S. Abu Issa, A. & Marsh, J. (2005). Reactions to Inspectable Learner Models: Seven Year Olds to University Students. *AIED'05 Workshop on Learner Modelling for Reflection, to Support Learner Control, Metacognition and Improved Communication between Teachers and Learners*. 23-32

[19] Zapata-Rivera, J.D., & Greer, J. (2004) *Interacting with Bayesian Student Models. International Journal of Artificial Intelligence in Education.* Vol. 14, Nr. 2 127-163.