metadata, citation and similar papers at core.ac.uk
brought to you by
provided by Kwansei Gakuin Univer

Language Acquisition Database

David WRIGHT（Language Center）
Toshiyuki TAKEDA
（Center for Research into and Promotion of Higher Education）

ABSTRACT

The application of Learning Analytics to oral communication has not been adequately addressed in Applied Linguistics literature. This has resulted in a limited number of practical tools for use in higher education, English as a Foreign Language（EFL）, settings. This lacking can be addressed through the development of handheld database technologies, which will allow more complete assessment of a learner's oral language ability and also the provision of learner specific corrective feedback（CF）. In addition, long-term tracking will better expose the process of second language acquisition（SLA）. This paper focuses on the design of a learning analytics software application for use on a handheld device in practical classroom settings. A paper version of the database has been formatted through a paper prototyping activity. The pilot version contains several core elements related to EFL instruction. First, sort functions will allow ordering of student records by name（alphabetically）, several measures of language proficiency, and a random list. Language proficiency will have set criteria including: ability level, accuracy, and fluency. In addition, a function to take audio samples will allow analysis by independent raters（i.e. other instructors）to ensure inter-rater reliability and to allow corpus linguistic level data analysis. Reports based on the data will include statistical analysis of learner performance criteria, including: mean, median, and rank. The existing design and structure will facilitate the expedited development of the system for real world use in educational settings.

1. INTRODUCTION

The temporal nature of oral language production makes systematic analysis in practical EFL teaching settings difficult. Current approaches to the evaluation of oral communication have significant limitations. For example, paper-based analysis requires long turn-around times, as by necessity, the data is manually input and then analyzed by an instructor. Therefore, any action taken, such corrective feedback（CF）provided to a student or class, is delayed by more than 24

hours. The use of automated data analysis will allow immediate, in-class review of individual student production from a historical perspective, as well as, comparisons across groups, classes, and program-wide. Future iterations of the application could include partial replication of methods used in theoretical SLA (second language acquisition) research, such as error token counting, which will increase the reliability and validity of data collected in practical settings. It is also important to note that the current system has been designed with an acknowledgement that data will not meet the more precise standards set in more theoretical settings. In fact, the intended goal of such a database is to bring increased levels of systematic analysis and on-demand CF to practical settings, so as to bridge the gap between the practical and theoretical data analysis.

2. LITERATURE REVIEW

The required use of labor-intensive processes has historically presented a barrier to data analysis of oral communication in SLA research.

2.1 Limits to implementation

Data collection of oral communication production often involves audio recording and subsequent transcription, the content of which is then analyzed by hand, including tabulations of set criteria. The development of software tools, such as grammatical parsing applications, has aided this process, but also require time intensive data processing. In response, EFL instructors have developed alternate ways to analyze student output, particularly observation, in either an interview format or through direct observation of participants engaged in a communicative activity. Data is collected on paper, in the form of language samples, subjective criteria rankings or objective measures, such as error token counting.

Results are then transferred to an electronic format for analysis. The data can then be analyzed in parallel with data collected from other sources, such as internet based assignments and tests. Hunter (2011) provides a working example of a database that provides authentic language samples for review within 24 hours. The assignments are tailored for both individual errors and class level common errors. The study provides a significant improvement in the analysis of oral production. However, the lag time does not allow immediate feedback on oral production, in the same way that it is immediately provided from computer aided and internet-based activities and tests. This gap, in the timing of informed feedback, will be addressed by the database described in this paper. The database will also plug-in to existing technologies.

2.2 Learning Analytics in Education

Learning analytics tools have been implemented at the institutional level, via integration into Learning Management Systems (LMS). This has allowed tracking of web based testing as well as attendance. An LMS system acts as a central point for data collection, and can be configured with

dashboards, that display data in various forms and in so doing, have significant potential to transform education for both the instructor and the student (Ferguson, 2012). This represents a new process, using a new paradigm noted as 'technology enhanced learning' (TEL), which pushes educators to become instructor-analysts, merging education methodologies with technology tools.

Further, as predictive algorithms improve, the potential to individually tailor outcomes for students increases (Takeda, 2012). It is within such a collection of student content, that tracking and analysis of oral production data is needed. The application of learning analytics tools to classroom activities will allow individual student and class specific feedback to be provided within the same class period. Though such a system has been theoretically possible using personal computing devices, the stationary nature of such systems limits interaction, as the instructor is often left in a fixed position or made to carry unwieldy equipment. The recent advent of wireless tablet computing and database applications, for example, an iPad linked to Filemaker Pro Server, allows the synthesis of traditional paper-based observation with immediate data analysis of student production. The results can be employed during the same class to structure activities around curricular goals and research methodologies. This will be particularly effective in providing immediate corrective feedback to individual learners.

2.3 Enhanced Corrective Feedback

Theoretical research by Sheen (2006; 2010) has demonstrated the effectiveness of focused recasts, a form of oral CF, where a single target structure is targeted and an instructor highlights errors immediately. However, the selection of target structures in mixed-level classrooms becomes problematic. A system of tracking individual student errors (i.e. accuracy) will allow focused recasts on specific area of improvement for each student, via sorting of individual student records within the database. Further, grouping students by differing ability level, with an emphasis on peer-to-peer instruction, or by matching students of similar levels, the system will allow a variety of activity sets within one class, with each activity targeting the specific needs of each student. In addition, feedback specific to each group can also be provided, allowing collaborative learning to be fostered. Such a capability, when combined with the data from various sources (e.g. the database created in Hunter 2011) and integrated into an LMS system (such as the LUNA LMS system at Kwansei Gakuin University), will create the potential for a unified dashboard, noted by Ferguson (2012), that better informs both the output and subsequent outcomes of language learning through the employment of tailored CF. In addition, systems such as ePortfolio could be added, to allow personalized displays of student generated content, for use in academic and professional development. This will address the SLA construct of individual differences (Dörnyei & Skehan, 2003), by allowing students to receive both instruction in a manner most suitable for their sensitivities and feedback based on their specific language needs. Achieving this within a mixed level EFL course represents a significant improvement in the efficacy of EFL teaching, as it exists at present.

3. RESEARCH PROPOSAL

The authors of the current study propose the creation of a Language Acquisition Database (LAD) to enhance the teaching of oral communication in higher education EFL settings. The initial design will include tables tracking language ability, accuracy and fluency. These will be related to student and class tables to allow filtering by student or class. The design will be created in-house at Kwansei Gakuin University (KGU) through collaboration between faculty and students. The process will simulate real world software development procedures giving practical experience to the students involved. Once the initial database is developed, the user interface (UI) will be pilot tested on campus at KGU.

3.1 Method

The LAD system will be validated through pilot testing and peer review. KGU instructors will be recruited to test and use the system in real-world classrooms settings. Feedback from such experiences will inform the further development of the system. Initially, the design will mirror common constructs in SLA research and EFL teaching methodologies, mentioned above. The primary function of the system will be to provide real-time data analysis of leaner oral production.

3.2 Design

The initial performance measures will include language ability, accuracy, and fluency. Each measure will have a form on the database, which will be accessible from a tab at the top of each screen. Each form will also have a 'record' button to allow for audio recording, that will be time stamped and synchronized with the markings by the instructor. This will allow independent raters to evaluate student performance. This will be primarily used for to establish inter-rater reliability, but could also be used in entry/exit proficiency interviews for the purpose of providing reliable and valid grades for student oral proficiency as well as research projects. Further, each language proficiency measure will have other distinct features that relate to specific SLA constructs.

3.2.1 Language Ability

Language ability will include five criteria: grammar, fluency, listening ability, text type, and pronunciation (see Appendix A). The items will be measured on a subjective score, using a Likert-scale from 1 to 5. Grammar will measure *accuracy*, including errors such as subject verb agreement, *continuity*, regarding the speaker's context and that perceived by the interlocutor, and any *target grammar* from a given lesson or activity. Fluency will measure comprehensibility, gestures, and naturalness of speaking. Listening will include appropriate responses to questions as well as reactions to various statements. Text type will measure the complexity of utterances, for example, the use of clauses. Pronunciation will include the phonemic level, stress patterns and rhythm. Portions of the above will also be analyzed using other criteria. However, the intention of

the language ability criteria is to provide a snapshot of language proficiency in a manner that allows the instructor to draw on subjective analysis, including experience and expertise with EFL students. Reports based on the data will allow general feedback, which can be provided regarding areas of strength and weakness. These will then be immediately linked to classroom or homework assignments.

3.2.2 Accuracy

The second criterion, accuracy, represents a more objective measure of performance (see Appendix B). It will be measured by error counting, through the use of counter buttons (also known as 'steppers') on the database form. Errors to be counted will include grammar, question and answer (Q&A) interactions, vocabulary, pronunciation, and target items from a given lesson. Grammar will include syntactical, grammatical and continuity related errors, as mentioned above. Q&A will count pace and timing errors during interactions, matching question forms with answer forms, and comprehensibility among participants and the instructor, who will act as an observer. Vocabulary will concentrate on academic terminology, collocations and lexical items, especially word choice. Pronunciation will be marked for syllable stress within words as well as sentence level stress and rhythm. Target items will be based on lesson content. Error rates will be calculated based on the number of errors over a given period of time.

A timer will be initiated when the data entry process begins on the form, allowing a calculation to be executed when the operation is completed, using fields labeled "date created" and "date modified". The result will show the number of errors per minute for each session. This will allow for comparisons across the student population, regardless of how long each student is monitored. However, a secondary entry will provide how much time has been spent with each student. This will allow the instructor to balance attention across the class. Most critically, this function will be used to identify student specific errors, including which type of errors each student generates over time. From the data, various reports will allow analysis of historical errors, as well as those made more recently, including the current session. This will allow CF, such as focused recasts during a class. A similar approach to CF will be addressed in the final criteria, again through the use of steppers to count anomalies.

3.2.3 Fluency

Fluency will be measured by the following criteria: pronunciation, pace, pauses, filler, initiate/extend, vocabulary, and target language (see Appendix C). The counter will be used to mark deviations from each category. For *fluency*, pronunciation will only include general comprehensibility. *Pace* will track the speed of language production, if it is unnaturally slow or fast, it will be marked. *Pauses* used to consider a point or allow an interlocutor to respond will be considered appropriate, whereas long pauses in replying, without the use of fillers, or other extended periods of silence, will be marked. *Fillers* such as 'umm' will not be marked, unless a single item is over used or used in an unnatural context. During the flow of a conversation, the ability to *initiate/extend* a conversation will be examined, for example if there is an abrupt change,

without explanation, an entry will be marked, further, if during a Q&A or other context, additional information is not offered voluntarily, an entry will be marked. *Vocabulary* will be marked according to the context, for example, in that casual expressions are not appropriate in business contexts. Finally, as with previous categories, *target* usage will be measured and marked by how naturally it is handled by each student.

3.3 Reporting

The above data will be presented in a simple report showing the mean and median for each criterion within each category. The mean of each criterion will be ranked to indicate which area is most problematic, with the rank of 1 representing the most frequent occurrence of an error or incorrect usage. The three categories will also be ranked. Both rankings will allow leaner specific feedback based on the individual differences of each student. Further, the overlap of content in level ability, accuracy, and fluency will allow instructors to refine their grading schemes, by comparing their own subjective evaluations with more objective counts. However, this aspect of the design will not be addressed in the current paper.

3.4 Participants

KGU faculty, staff, and students will contribute to the design, development, and implementation of the LAD system. KGU students and staff will be recruited as participants in follow-on studies. The participant pools will be expanded, for example to external institutions, based on the results for initial pilot tests.

3.5 Results

In addition to the publication of the current white paper, additional publications will be generated throughout the development process. Interdepartmental and Institutional research will follow with an eventual focus on cooperation with other academic institutions, including those in ESL settings. The results of which will contribute to several fields, including: EFL teaching methodologies, Informatics, Learning Analytics, and SLA research.

4. CONCLUSIONS

The prototyping process has yielded a workable database design that, once in operation, will contribute to EFL classroom pedagogy and SLA research. The limited complexity of the initial system will foster rapid development. The lag in providing immediate CF is a technological obstacle that can be addressed by the database as described above, which will then contribute to the academic literature on the subject. This solution will provide EFL instructors at higher education institutions as well academic researchers with a new tool for measuring oral language proficiency in real-world settings. This will not simply compliment theoretical studies, by bringing adapted methodologies into in-tact classrooms, it will also foster inter-departmental

cooperation among empirical research and practical research by providing corpus data sources for empirical research. This will build cooperation and facilitate the process of bringing advanced research concepts into EFL classrooms in an expedited manner.

4.1 Limitations and future research

The original proposal conceived by Instructor Wright envisioned a complex database that included multiple items to track student data and complex algorithms to analyze that data, including data from written sources. The prototyping process revealed that the scale of such a project required that it be sub-divided into smaller elements, the first of which being the design noted above. The experience will also inform the further development of the system, allowing for the completion of elements in manageable units. For example, future iterations will employ customized performance criteria for specific EFL classrooms, as well as, predictive algorithms to guide learning outcomes. In addition, reports will become more detailed and included graphical representations of data (i.e. pie charts) that better highlight the language learning needs of each student.

4.2 User interface development

The user interface (UI) will be refined and additional elements will be added. The long-term goal is to create a language proficiency diagnostic tool that can be used to predict student communicative competence in all four modalities (listening, reading, speaking, writing). The sum of these elements are to be integrated into the LUNA LMS system, via plug-ins, to create a seamless user experience for faculty and students, with a particular emphasis on customizable dashboards for both students and instructors. Funding will be requested, from internal and external sources (i.e. including 'Kakenhi' grants such as Scientific Research on Innovative areas) at various stages of the development in tandem with updates and published reports regarding the development process. Future iterations will include more specific criterion and be integrated into systems such as an ePortfolio.

4.3 Expanded evaluation criteria

The current design has a general category for grammar, however, if resources are available, this will be expanded to identify specific tense/aspect system level errors (Celce-Murcia & Larsen-Freeman, 1999). Most simply, this could include past, present, and future distinctions. The long-term goal would be to include counters for each of the 12-tense/aspect items, so as to specifically inform students regarding their grammatical accuracy. In the interest of an expedited development, this sub-division will remain a later add-on. Further, future research efforts will examine whether recasts of historical or more limited (recent) records are more effective when applied to intact classrooms. From this, the database will be expanded to include other criteria, such as developmental stages (i.e. Pienemann 1998), student motivation, and communicative

TBLT task (Ellis, 2003) performance. Data collection will also be expanded to include video capture for analysis of non-verbal communication and research into conversational interactions (e.g. Mackey, 2007). There is also the potential, to provide students with hand-held devices. The creation of data entry forms, and the pairing of devices, would allow students themselves to enter data based on their interactions, further expanding the data set. These represent a limited number of possibilities, many more of which will likely be presented in the feedback from KGU instructors after piloting the system. With this in mind, a modular system will be created that will allow future add-ons to be implemented rapidly.

Special Thanks:

The researchers would like to express their gratitude to the members of the Learning Analytics Project, who provided invaluable feedback during the prototyping process. We would also like to thank the administration and faculty at our home institution, Kwansei Gakuin University, for their support of academic research and professional development.

References

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The grammar book: an ESL/EFL teacher's course* (2nd ed.). Boston: Heinle & Heinle.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition*. Oxford, England: Blackwell.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, England: Oxford University Press.

Ferguson, R. (2012). The State Of Learning Analytics in 2012: A Review and Future Challenges. *Technical Report KMI-12-01, Knowledge Media Institute, The Open University, UK, 1*. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Hunter, J. (2012). 'Small talk': developing fluency, accuracy, and complexity in speaking. ELT Journal, 66(1), 30–41.

Mackey, A. (Ed.). (2007). *Conversational interaction in second language acquisition*. Oxford, England: Oxford University Press.

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Philadelphia, PA: John Benjamins.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, 10(4), 361–392.

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32(2), 203–234. Doi: 10.1017/S0272263109990507

Takeda, T. (2012). Notes on Academic Analytics in Higher Education. *Kwansei Gakuin University Researches in Higher Education*, 2, 37–48.

APPENDIX A

Data Entry - Ability (D3a)

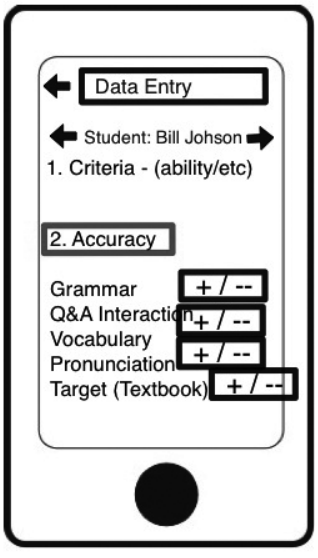
Data parameters (for all data entry, tasks, assignments, AV recordings)
1. Count total correct & errors, for calculating ratio
2. Calculate time each student page is open to determine time spent on each student
3. Record date and time or data entry, for use in algorithms to calculate frequency of errors, level up/down, & overall language proficiency

APPENDIX B

Data Entry - Accuracy (D3b)

Data Input

1. Ability
2. Accuracy
3. Fluency
4. Devel Stages
5. Level Check
6. Custom Criteria
7. Writing
8. Reading
9. Motivation



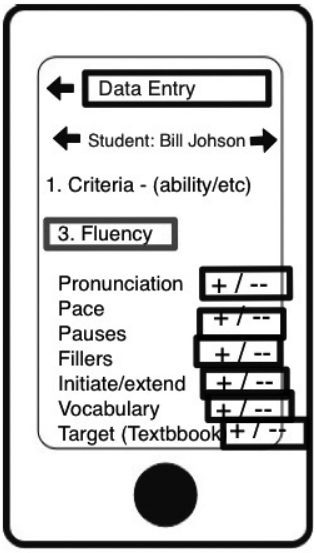
Data parameters (for all data entry, tasks, assignments, AV recordings)
1. Count total correct & errors, for calculating ratio
2. Calculate time each student page is open to determine time spent on each student
3. Record date and time or data entry, for use in algorithms to calculate frequency of errors, level up/down, & overall language proficiency

APPENDIX C

Data Entry - Fluency (D3c)

Data Input

1. Ability
2. Accuracy
3. Fluency
4. Devel Stages
5. Level Check
6. Custom Criteria
7. Writing
8. Reading
9. Motivation



Data parameters (for all data entry, tasks, assignments, AV recordings)
1. Count total correct & errors, for calculating ratio
2. Calculate time each student page is open to determine time spent on each student
3. Record date and time or data entry, for use in algorithms to calculate frequency of errors, level up/down, & overall language proficiency