Building a digitized oral proficiency rubric to enhance Corrective Feedback

David Wright

Building a digitized oral proficiency rubric to enhance Corrective Feedback

David Wright

1. Introduction

The ability to provide corrective feedback (CF) in a variety of formats, such as presenting the results of a speaking test by using an oral proficiency rubric, is an established component of second language instruction, particularly at universities in Japan. Building on this, hand-held computer technology allows for additional forms of enhancement and dissemination of rubric data, including automatic color-coding (e.g. highlighting) of scores and the ability to present such feedback in an informal situation, for example where an instructor casually walks around a classroom and gives immediate feedback based on in-class activities, that can optionally be evaluated for a grade. At present, this capability can be constructed on database software using a computer and then deployed to tablet devices. The use of a digital rubric allows for the visual enhancement of rubric data and increases the accessibility of the data, in class, via email, and online, while also helping support differing learning styles. In addition, it creates more opportunities to provide corrective feedback, and increases the range of instructional interventions, as there is more information (e.g. data) from which to evaluate individual, class, and course needs.

In this way, the concept of a needs analysis can then be integrated into teaching settings as a continuous process of evaluating, analyzing, and adjusting coursework (e.g. lesson goals) to student needs. The ever-expanding set of information related to student language proficiency that is collected will better validate understanding of students' language proficiency. This is particularly true when historical performance, relating to all students throughout the life of a

program, and related to all forms of assessment (e.g. norm-referenced tests, written work, and oral proficiency), can be evaluated in an aggregate form.

This in turn, when integrated into an overall curriculum, can promote self-directed learning through learner review and analysis of artifacts (e.g. an email of a rubric score) and interactive online activities. Going further, critical thinking skills can be developed as students are tasked with creating their own ways to push the development of their own language proficiency based on a regular flow of feedback. Related to that, from a practical perspective, such a process may allow more class time for higher-order thinking skills, as students can provide peer-feedback (assuming the presence of a class set of tablets or the use of personal smartphones), as well as self-analysis of recorded samples of learners' own production.

In the current paper, the author will describe a database, developed using FileMaker Pro software and deployed on an iOS device (i.e. an iPad) using the FileMaker GO app. The system represents a digitized iteration of a paper-based oral proficiency rubric that is in use at the author's home institution. The design is similar to many learner management systems (LMS), in that students can be assigned to many classes and evaluated within each class. However, a core difference is that the software runs natively on an iPad and can optionally be connected to a server (which would essentially mimic an LMS). The project began in reaction to the need to provide LMS-like functionality without the need for a wireless connection to a server, using an app running natively on a tablet device. Given that, a primary goal of the project is to fill a gap in the research regarding the development of LMS-like software (e.g. a language learning database) that can operate without an internet connection. The project began several years ago, with the creation of a paper-prototype (i.e. Wright & Takeda, 2013), then a working version using simple bar graphs and numerical scores, and finally the version explained in depth here, that represents a touch-based, interactive version of a paper-based oral proficiency rubric. The paper will begin with a theoretical rationale for the database, including a brief literature review of studies related to the project.

2. Rationale

The provision of corrective feedback is more effective when it is given in such a way that it is most likely to get the attention of the learner, so as to prompt repair in the learner's interlanguage, and to prompt more fluent utterances in future conversational interactions, which is what Swain (1985) refers to as uptake. The current paper draws on the concept of uptake, by analyzing the presentation of a specialized form of textual-enhancement (Han, Park, and Comb, 2008) to verify if rubric data is more salient to English Language Learners (ELL) when color-coded. Such a process automates the tabulation of scores, which builds on traditional forms of paper-based oral proficiency rubric evaluation, by allowing for immediate debriefing after the completion of an assessment. In the author's experience, a paper rubric is annotated (e.g. highlighted in one color) by either the instructor or a student and the results are then discussed with the student. The paper then becomes a static artifact, showing the student's proficiency. In contrast, a digital version of the document can be designed to have greater visual impact, through the use of multiple colors and touch functionality. In addition, it could be distributed to the student as a form of digital artifact, to mark their progress and for use in activities based on the results. The process becomes a digitally enhanced form of delayed corrective feedback, where portable tablet computers allow for novel ways to enhance the assessment of rubric content, as well as new and persistent (e.g. online) ways to present the information.

The digitization of content also allows for new practical solutions to promote self-directed learning. The capability to record audio and video content allows learners to evaluate their own oral production and that of their peers, by using a digital rubric, which is in essence, a form of stimulated recall (e.g. Gass & Mackey, 2000). This can be particularly effective when the digital rubric is linked to an online access point, which is a future goal of the current research study, to allow self-reflection outside of the classroom. Persistent access to one's own data provides the ability to self-reflect on one's own language production, which then

allows for learning opportunities, for example to integrate the data into curricula, such as CLIL, or content and language integrated learning (e.g. de Zarobe, 2017), where content and language are taught simultaneously. In such a case, the data could be expanded to include tracking of content knowledge and intercultural communication skills, alongside language proficiency. With the above in mind, the design of a digitized oral proficiency rubric will be described in detail.

3. Database Design

The database was created with great consideration to providing a novel form textual-enhancement (e.g. Han, et. al. 2008), specifically automatic color coding of the written elements of the rubric, each of which is based on a range of numerical scores. First, internally, a relational database was created with an initial relationship graph that allowed for many-to-many relationships, for example a 'Students' table and a 'Classes' table were linked using a 'Sign-ups' join table, with the Key fields for both source tables being assigned as Foreign Key fields in the 'Sign-ups' table, so that each student could be in many classes and each class could have many students. This was also done for an 'Audio/Video Samples' table, which would be used to collect audio and video samples of language production from within the App. Building on the initial set of relationships, the 'Audio/Video Samples' table was joined to the 'Sign-ups' table using an 'AV Sign-ups JOIN' table. However, to allow for participants in different classes to by recorded together, a second iteration of the relational graph was created, where the 'AV Sign-ups DATA' table was linked by the 'Classes' Key field (via the Foreign Key in the new 'Sign-ups VALUES' table). This simplified data entry in that selection of a specific class would limit the list of students to only those in the current class and would allow ad-hoc combinations of students from different classes, for example during a make-up test. Once the internal design was established, a set of layouts were created to act as the user interface, with the primary interface being the digitized oral proficiency rubric.

The working database was then pilot tested, as part of a larger Japan Society

for the Promotion of Science grant, using a set of iPad mini 2's, with data collected from university EFL instructors. A survey was conducted regarding the ease of use of the database and how to improve it. The survey results showed that the user interface was fairly intuitive, but would still require familiarization training (Wright & Nakagawa, 2016), with a set of specific requests regarding the location of specific buttons. One critical comment made by the instructors in the survey was the limited size of the 7.9-inch screen on the iPad mini. The comments related to the small text size when displayed on the screen, which made reading specific rubric boxes challenging, during a feedback session. This form of eye strain, when added to the pressure on an instructor to accurately rate a learner, prompted the need to reconsider the size of the device screen. Anecdotally, the author also wondered if difficulty reading the feedback would also add to a participant's (i.e. students) cognitive load when being tasked to reflect on their foreign language proficiency, while communicating in a foreign language. To improve the readability (i.e. text size) and the visual impact (i.e. size of the color-coded rubric itself) the current paper proposes the use of an iPad Pro, 12.9-inch version, while utilizing the same layout format, expanded to fit the larger screen. A device was procured using funds from the Centre Project Fund of Gakushuin University. The revised layout is in general more user friendly, based on changes made in reaction to the instructor survey, and the version for the larger iPad Pro features text sizes that are easier to read, as shown in figure 1.

Figure 1: Textually Enhanced Oral Proficiency Rubric.

As shown in Figure 1, the layout for the oral proficiency rubric is essentially an interactive paper rubric. There are invisible buttons above the rubric content grid squares, with two buttons for each square, that are labelled with a number. When a rater touches one of the buttons related to a specific grid space on the rubric, a score is automatically input and shown above the column relating to that category. There is also the option to directly input a number by touching the number score box itself. Once a score is entered, the column changes, with a color-

coded scheme. In the scheme, the current score and those below are highlighted in yellow, while the next score up (i.e. the next target in a learner's development) is in green. This provides a 2-D bar graph with color-coded textual enhancement (e.g. Han, et. al. 2008) encapsulated within it, so as to address the results of an earlier study (Wright, 2015), where participants found little relevance to a traditional bar graph. The students' attention is directed to the green box, to focus on how to get to the next level. In addition, the language displayed can be switched, with currently 2 possible language combinations, to ensure the participants can fully comprehend the content of the rubrics in their native language.

Several other layouts create the base for a quasi-learner management system (LMS) that runs natively on the iPad, and includes the ability to record language samples and easily navigate to different students and to differing classes or treatment groups. The entire database was built by the author, through a series of approximately 100 design iterations over a period of 3 years. In regard to the current paper, a working version of the system was recently pilot tested by the author at a university in Japan, using a 12.7-inch iPad Pro.

4. Practical Study

Two groups were recruited for the study, one (Group A) from students taking oral communication courses and another (Group B) from students taking reading and writing courses, with volunteers from each group signing consent forms prior to the start of the research. During the treatment period, the participants in Group A were enrolled in an oral communication course, where they engaged in regular conversational interactions both among the students and between individual students and their classroom teacher. The level of the students ranged from upper intermediate to advanced. Three students opted to participate in the study (n=3). Participants in Group B received minimal instruction related to oral communication during the treatment period, as they were enrolled in English reading and writing courses. Initially, 12 students volunteered. However due to practical limitations, the researcher was able to collect a full set of data from nine students (n=9). The

students in the B group were evaluated as beginners based on a listening test. The researcher conducted the data collection at times mutually agreed upon by the students and the researcher. The participants were unfamiliar with the researcher prior to the research. Both groups were given a pre-test, where they engaged in a short conversation with the researcher, and the conversation was recorded. They were then provided feedback over a 6-week period, with individual students receiving from 3 to 5 feedback (treatment) sessions. Each participant in Group A received 5 feedback (treatment) sessions. The members of the B Group received 3-4 feedback (treatment) sessions. The treatments were followed by a post-test, where another prompted conversation took place with the researcher, with the event also being recorded. The B group was given the post-test in a somewhat rushed manner due to scheduling conflicts. The recorded testing measures were evaluated by an independent rater, using the language acquisition database described in this paper. Practical constraints meant that creating a control group was not possible. In addition, given the small sample size (N=12), with an unbalanced number of proficiency levels, the results were not examined for significance, given that any form of generalization related to a larger cohort of learners was not possible.

The results of individual scores varied, with participants generally maintaining their ability across the experiment, with the average of the total score, out of 50, for Group A decreasing 1.3 points, whereas the average scores for the B Group decreased by 1.4 points. In terms of the B Group, the lack of noticeable changes (e.g. a dramatic difference between the testing measures) were likely due to the short length of data collection and the low number of treatments. The limited number of feedback sessions in the second group was a result of difficulties arranging an appropriate time for the researcher to meet the participants. In addition, during the early sessions, the researcher was still building rapport with many of the students in the study. This, along with the small number of corrective feedback interventions and the conditions of the posttest, where there was a need to expedite the testing measures, likely influenced the scores. On a practical level, this indicates that it would be beneficial to have a course instructor conduct the feedback sessions or to

have an outside researcher interact with the participants several times prior to the recording of the testing measures if data is collected from intact classes. However, there were no signs of improvement in the A group either. This may have been due to the limited number of interventions, it may also indicate the need for adding a task after each treatment, as a form of self-reflection, or in a classroom context, an assignment. For future studies, there is clearly a need to increase the number of treatments, for example over the course of a semester, and have the instructor of the course conduct the feedback. Another issue that arose during the experiment was the functionality of the database user interface.

The database developed for this project includes a user interface (UI) custom designed by the researcher. The main page allows a single participant to be assessed and the results displayed by scrolling down a screen using a swiping gesture on the iPad. However, there is only a limited function to search for a specific student, and no function to evaluate two students at the same time. These issues were identified early in the development process, but were beyond the scope of the original design timeline. The researcher is currently working to develop a database with such functionality. Building on the need for added functionality, a secondary goal is to link the system to an online platform.

The FileMaker Pro database can itself be hosted, however it can also be linked to other online learner management systems (LMS), either in real-time or through an a-synchronous importing and exporting of data. This is a goal that the project related to the current paper is focusing on. The primary reason is that this will allow a new data set for the LMS to utilize, for both evaluation purposes and in other new ways. For example, oral production samples could be collected in a classroom and then evaluated by students and their peers as a post-task assignment, which would help learners attend to their production in a more informal manner, that is without the presence of an instructor. A follow-up study will seek to examine this process in an experimental setting, and if possible classroom settings during later iterations.

6. Conclusion

Individualized corrective feedback provides learners with detailed information on their skill level and proficiency. When learning a language, feedback on oral production has been traditionally limited by the nature of paper-based evaluation methods. The digitized oral proficiency rubric presented here, the second iteration of a functional database design and the associated user interface, both developed by the author, represents a tool that language instructors can use to provide multiple evaluations with automated tabulation and the added value of textual enhancement. However, the introduction of the technology itself, as well as practical planning (e.g. scheduling) require careful introduction to the participants and to the instructors who will collect the data, to ensure that they buy in to the value the system will provide. The results of a small-scale practical study demonstrate that a small number of feedback sessions was not enough to impact language proficiency. Given that, future studies will include longer treatment sessions, as well as a formal introduction of the database, to ensure this form of corrective feedback provides the metalinguistic push needed to prompt improvements in oral proficiency.

Acknowledgments

References

de Zarobe, Y. R. (2017). Language Awareness and CLIL. *Language Awareness and Multilingualism*, 149-159.

Gass, S. M., & Mackey, A. (2000). Stimulated recall methodology in second language research. Routledge.

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4), 597-618.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), Input in second language acquisition (pp. 235–253). Rowley, MA: Newbury House.

Wright, D. (2015). The Effects of Visual and Text-based Feedback on Oral Proficiency Assessment Rubrics. *The Bulletin of the Foreign Language Center, Tokai University*, 35, 1-15.

Wright, D. & Nakagawa, H. (2016). Language Acquisition Database: user interface needs analysis. Paper presented at the 2016 International Symposium on Language, Linguistics, Literature and Education, Naha, Japan, Aug. 4.

Wright, D., & Takeda, T. (2013). Language Acquisition Database. *Kwansei Gakuin University Higher Education Research*, 3, 75-84.

修正フィードバックの効果を高めるための
デジタル版口頭運用能力評価基準の開発と運用

デイビッド・ライト

　外国語の授業におけるスピーチやプレゼンテーションなどの発表活動においては、活動中やその直後に様々な観点からの詳しい評価が行われ、その結果とともに何をどう改善すべきか等の修正フィードバックをすぐに得ることができれば、形成的評価として最も効果的であると考えられる。さらに、それらが記録として蓄積され、それを後で参照することで学習者が自分の活動をいつでも「振り返る」ことができることが望ましい。そこで、本研究では発表活動について、詳細な評価を教員が iPad 上で短時間で行える口頭能力評価基準データベースを開発した。具体的には学習者の活動中に教員が iPad を用いてそのパフォーマンスを記録および評価し、それを活動直後に学習者に修正フィードバックとして提示したり、評価データをサーバーに送ってそこに蓄積することにより、後で学習者が手元のスマートフォン等で参照できるようにするとともに、時系列で向上の度合いがわかるようにした。

　本稿ではこの実践研究についての意義と口頭能力に関する評価基準データベースの開発について詳述するとともに、本研究の大人数を対象とした本格的な実施を前に、英語を外国語として学ぶ日本人大学生 12 人からなる実験群に対して行った実践について述べる。得られたデータが限られているため、その効果について一般化することは現時点では難しいが、こうしたシステムを実際の授業で活用することで、発表活動が行われている中でも教員による詳細な評価活動が可能となり、学習者は活動直後やしばらくした後、その評価結果やフィードバックをスマートフォン等で参照することで、評価をその後の活動に役立てられることがわかった。