Self-Assessment of Motivation: Explicit and Implicit Indicators in L2 Vocabulary Learning

Kevin Dela Rosa, Maxine Eskenazi

Language Technologies Institute, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, Pennsylvannia, USA
`{kdelaros,max}@cs.cmu.edu`

Abstract. Self-assessment motivation questionnaires have been used in classrooms yet many researchers find only a weak correlation between answers to these questions and learning. In this paper we postulate that more direct questions may measure motivation better, and they may also be better correlated with learning. In an eight week study with ESL students learning vocabulary in the REAP reading tutor, we administered two types of self-assessment questions and recorded indirect measures of motivation to see which factors correlated well with learning. Our results showed that some user actions, such as dictionary look up frequency and number of times a word is listened to, correlate well with self-assesment motivation questions as well as with how well a student performs on the task. We also found that using more direct self-assesment questions, as opposed to general ones, was more effective in predicting how well a student is learning.

Key words: Motivation Modelling, Intelligent Tutoring Systems, Computer Assisted Language Learning, Motivation Diagnosis, English as a Second Language

1   Introduction

Motivation modelling and its relation to user behavior has receieved attention by the educational computing community in recent years. William and Burden define motivation as "*a state of cognitive and emotional arousal which leads to a conscious decision to act, and which gives rise to a period of sustained intellectual and/or physical effort in order to attain a previously set goal (or goals)*" [1]. The use of self-assessment questionaires is one common approach to measuring motivation. One such construct is Motivated Strategies for Learning Questionnaire (MSLQ), an 81-item survey designed to measure college students' motivational orientations and their use of various learning strategies [2]. While questionnaires are useful to detect enduring motivational traits, some are criticized, particularly those administered prior to interaction. Since a student's motivation is likely to change during an interaction, it is important to use them with other methods to adapt instruction and to gather more transient information about a student's motivation [3]. Other methods of assessing

motivation include direct communication with students, emotion detection, and recorded interactions with an intelligent tutor. For modelling and understanding user behavior automatically, Baker [4] showed that machine learning models trained on log data of student activity can be used to automatically detect when a student is off-task. A study by Cetintas et al. [5] reached a similar conclusion, using a regression model personalized to each student. And, Baker et al. [6] showed that a latent response model can be used to determine if a student is "gaming" the system in a way that leads to poor learning.

An important issue has been how to automatically detect a student's current motivational state. As mentioned above, one method of measuring motivation is questionnaires that cover a variety of motivation aspects. One important consideration is how detailed and/or direct these survey questions should be with respect to the task or, in other words, is it better to have questions that are tightly focused on the tasks being performed by the student or is it better to construct questions that are more general and can cover many difference aspects of motivation, such as the MLSQ. Also, a student's usage of a tutoring system, as indicated by the amount of activity and types of actions taken, may furnish good implicit indicators of a student's motivation.

We propose that in a computer-assisted L2 language learning environment certain recorded student interactions during learning activities can act as implicit indicators of that student's motivation. We also propose that these implicit indicators, as well as explicit ones like self-assessment surveys, can be used to predict the amout of learning that is taking place. Lastly we postulate that more *direct* questions may measure motivation better and may also be better correlated with learning.

For this study we used a web-based language tutor called REAP [7]. REAP, which stands for **REA**der-specific **P**ractice, is a reading and vocabulary tutor targeted at ESL students, developed at Carnegie Mellon University, which uses documents harvested from the internet for vocabulary learning. REAP's interface has several features that help to enhance student learning. One key feature in REAP is that it provides users with the ability to listen to the spoken version of any word that appears in a reading, making use of Cepstral Text-to-Speech[1] to synthesize words on demand when they are clicked on. Additionally, students look up the definition of any of the words, during readings, using a built-in electronic dictionary. REAP also automatically highlights focus words, the words targeted for vocabulary acquisition in a particular reading. REAP is a language tutor and a testing platform for cognitive science studies [8, 9], as is the case of this study.

In this paper we describe a classroom study that compares the effectiveness of different motivational indicators in a vocabulary learning environment. We define the different types of survey questions we used as explicit measures of motivation and the various user actions we recorded as indirect indicators of motivation. Next we describe the results of a classroom study that integrated our various motivation indicators and how well they correlated with our learning measures. Finally we discuss the implications of our results and suggest future directions.

---

[1] Cepstral Text-to-Speech. `http://www.cepstral.com`

2 Classroom Study

In order to determine which of our hypothesized indicators of motivation were most related to learning we conducted a classroom study with a web-based tutor, focused on L2 English vocabulary learning, and recorded responses to motivation questionnaires and user actions that we log, which may indirectly indicate a student's motivation level. The classroom study consisted of a pre-test and post-test with multiple choice fill-in-the-blank vocabulary questions, and six weekly readings, each followed by practice vocabulary questions similar to, but not the same as those in the pre-test and post-test. During the pre-test and post-test, a set of seventeen self-assessment motivation questions were administered, and after each weekly session there were a set of five motivation questions.

21 intermediate-level ESL college students at the University of Pittsburgh's English Language Institute participated in the study and completed all of the activities. For this study the readings and vocabulary questions had 18 focus words, taken from either the General Service List[2] or the Academic Word List[3], and not part of the class' core vocabulary list.

In the following subsections we describe the types of questionnaires administered, the recorded user actions, and the metrics we used to measure student learning.

2.1 Motivation Questionnaire

We administered motivation survey questions as explicit measures of motivation after each reading. 17 survey questions were administered in the pre-test and post-test, as shown in Table 1, using a five-point *Likert* scale, with a response of 5 indicating the greatest agreement with the statement and 1 indicating the least agreement. The 17 questions were divided into two groups: *General* and *Direct*.

We call *General* survey questions high-level survey questions which have been used in past REAP studies because of their generality; they are used in many studies in the Pittsburgh Science of Learning Center[4]. For example, one of the *General* questions we used was, "*When work was hard I either gave up or studied only the easy parts*", which can be used for many different subject matters. The design of these questions was guided by the MLSQ [2], and aimed to use the fewest number of questions possible that cover the most motivational constructs.

We call *Direct* questions the more explicit items that focused on aspects directly related to the reading activities accomplished over the course of the study. An example of a *Direct* question: "*Learning vocabulary in real documents is a worthwhile activity*". This is focused on the specific REAP tasks.

A total of twelve *General* and five *Direct* self-assessment motivation questions were administered during the pre-test and post-tests. Additionally, the five *Direct* motivation survey questions in Table 2 were asked after each weekly reading activity,

---

[2] The General Service List. `http://jbauman.com/gsl.html`

[3] The Academic Word List. `http://www.victoria.ac.nz/lals/resources/academicwordlist`

[4] Pittsburgh Science of Learning Center (PSLC). `http://www.learnlab.org`

at regular intervals in between the pre-test and post-test, to see how the responses correlated with student behavior and learning at each reading. We wanted to determine if there was a difference in how well each of these two question groups correlates to the learning measures we recorded (multiple choice questions). We hypothesize that questions more directly related to the tasks/activities performed will be better at predicting motivation and learning. This is guided by unpublished results of past REAP studies which have shown that higher-level questions generally failed to correlate well with learning measures, and previous success with direct questions by Heilman et al. [10].

Table 1. Pre-test/Post-test Motivation Survey Questions

ID	Survey Question Prompt	Group	Type
S1	I am sure I understood the ideas in the computer lab sessions.	General	E
S2	I am sure I did an excellent job on the tasks assigned for the computer lab sessions.	General	E
S3	I prefer work that is challenging so I can learn new things.	General	A
S4	I think I will be able to use what I learned in the computer lab sessions in my other classes.	General	V
S5	I think that what I learned in the computer lab sessions is useful for me to know.	General	V
S6	I asked myself questions to make sure I knew the material I had been studying.	General	O
S7	When work was hard I either gave up or studied only the easy parts.	General	A
S8	I find that when the teacher was talking I thought of other things and didn't really listen to what was being said.	General	A
S9	When I was reading a passage, I stopped once in a while and went over what I had read so far.	General	O
S10	I checked that my answers made sense before I said I was done.	General	O
S11	I did the computer lab activities carefully.	General	E
S12	I found the computer lab activities difficult.	General	A
S13	I continued working on the computer lab activities outside the sessions.	Direct	A
S14	I did put a lot of effort into computer lab activities.	Direct	A
S15	I did well on the computer lab activities.	Direct	E
S16	I preferred readings where I could listen to the words in the document.	Direct	V
S17	Learning vocabulary in real documents is a worthwhile activity.	Direct	V

Table 2. Post-reading Survey Motivation Questions

ID	Survey Question Prompt	Type
Q1	Did you find the spoken versions of the word helpful while reading this document?	V
Q2	Do you find it easy to learn words when you read them in documents?	E
Q3	Did you find this document interesting?	V
Q4	Did you learn something from this document?	V
Q5	Does reading this document make you want to read more documents?	A

Furthermore, for this study we grouped the questions into three types:

- **Affective** (*A*): Deal with *emotional reactions* to a task
- **Expectancy** (*E*): Deal with beliefs about a student's *ability to perform* a task
- **Value** (*V*): Deal with goals and beliefs about the *importance and interest* of a task

The 3 groups are based on the Pintrich and De Groot components of motivation (self-efficacy, intrinsic value, test anxiety) [11]. We used this grouping to simplify the analysis of the results. Note that in the tables and figures, "**Other** (*O*)" signifies a

question that failed to group into one of the three types, typically a question on learning strategies.

2.2   Recorded User Interactions

In addition to survey questions, we recorded actions taken by the students which we hypothesize would indirectly correspond to motivation and might also correlate with learning. The following were recorded during each activity:

> Word lookup activity, using our built-in electronic dictionary
> > *A1*: Total number of dictionary lookups
> > *A2*: Number of focus words looked up in the dictionary
> > *A3*: Number of dictionary lookups involving focus words
> Words listening activity, using our built-in speech synthesis
> > *A4*: Mean number of listens per word
> > *A5*: Total number of listens
> > *A6*: Number of words listened to
> Average time spent on activity tasks
> > *A7*: Time spent reading the documents
> > *A8*: Time spent on practice questions

2.3   Learning Measures

In order to assess how well students learned the target vocabulary words, we recorded the following measures:

> *L1*: Average post-reading practice question accuracy (for all questions appearing directly after reading the documents)
> *L2*: Pre-test to post-test normalized gain
> *L3*: Post-test accuracy
> *L4*: Average difference between pre-test and post-test scores

Note that that L2 and L4 are two different ways of looking at the improvements made by students over the course of the study, where L2 is tuned to the relative difference between the test scores, and L4 is sensitive to the absolute difference in scores.

3   Results

The results of our study show that the use of the REAP system significantly helped students improve their performance on the vocabulary tests, as evident in the average overall gains between the pre-test and post-test {*L3*} ($p < 0.004$), whose average scores were 0.3439 ($\pm$ 0.0365) and 0.5000 ($\pm$ 0.0426) respectively. The average post-reading practice question accuracy was 0.8417 {*L1*} ($\pm$ 0.0466). The overall average normalized gain {*L2*} between pre-test and post-test was 0.2564 ($\pm$ 0.0466), and average difference in score {*L4*} between the pre-test and post-test was 0.1561 ($\pm$ 0.0232).

Implicit Indicators	Implicit Indicators								Learning Factors			
	A1	A2	A3	A4	A5	A6	A7	A8	L1	L2	L3	L4
A1												
A2										-0.369	-0.497	
A3									-0.442	-0.421	-0.560	
A4										0.476		0.588
A5												0.490
A6												-0.395
A7												
A8												
Explicit Indicators												
Q1				-0.446	-0.530		-0.408	-0.535				-0.500
Q2		-0.402		0.553	0.458	0.391	0.526		-0.603		-0.389	
Q3				0.473			0.616		-0.501		-0.420	
Q4		0.507		-0.532			-0.374					-0.445
Q5		0.431		-0.474	-0.428							
QV				-0.625	-0.531			-0.410				-0.500
SA-General				0.546			0.635	0.415				
SE-General		-0.433	-0.502									
SV-General		0.405										
SO-General		0.645	0.454	-0.532							-0.413	-0.383
SA-Direct					-0.381					-0.437		
SE-Direct				-0.399	-0.431						0.399	
SV-Direct				-0.425	-0.455		-0.424		0.584		0.423	

Fig. 1. Significant correlations values between motivational & learning factors, and between implicit & explicit motivation indicators. Color signifies level of significance, with green representing strong statistical significance ($p < 0.05$), and yellow representing moderate signifance ($p < 0.1$). Note that self-correlations and correlations with low significance values were omitted. Also note that Q1-Q5 correspond to the students' average survey response values for all reading activities, and with respect to the implict indicators, A1-A8 correspond to the students' average values of those indicator values over all reading activties.

In order to find the motivational factors that best correlated with learning, we computed a Pearson correlation coefficient matrix for the values for the different factors and grouped question responses, and determined the significance of each pair of correlations using a two-tailed test. Figure 1 summarizes correlation and signifcance values found between the motivational and learning factors, and shows how the indirect indicators correlate with the explicit indicators of motivation.

Additionally, when we looked at the post-reading accuracies of each individual reading activity (as oppose to the overall averages, which were shown in Figure 1), the following motivational factors tended to significantly correlate with the post-reading practice question accuracies at significance levels varying levels over the six readings between $p < 0.01$ and $p < 0.05$:

– Q2 response: Do you find it easy to learn words when you read them in documents?

- Q3 response: Did you find this document interesting?
- Total number of dictionary lookups
- Number of focus words looked up in the dictionary
- Time spent on practice questions

4   Discussion

We see in Figure 1 that most of the *General* questions did not significantly correlate with the various learning measures, while the *Direct* questions did. In fact, the only sub-group of *General* questions that correlated with the learning measures was *Other*, those questions that did not fit well into our three types (*Affective*, *Expectancy*, and *Value*), which is not surprising since the *Other* questions mainly focused on learning strategies as opposed to motivation. Furthermore, all of the direct questions asked after each reading, except for *Q5* (*Does reading this document make you want to read more documents?*), had significant correlations with the learning measures. Perhaps the reason *Q5* failed to have significant correlations is due to the fact students were not given the option to actually act on the desire to read more documents in our tutor during the semester, due to class constraints. Therefore, our results imply that *General* questions are less effective at predicting the learning outcome of a student than are the *Direct* motivation questions which are more closely focused on the tasks performed. Moreover, our results hew closely to past results by Bandura [12] in the domain of self-efficacy.

Additionally, the implicit motivation indicators, based on recorded student actions, seemed to correlate well with our learning measures, particularly our word listening and dictionary lookup-related interactions, which implies that these kinds of actions can also help in predicting a student's motivational state. Interestingly, the amount of time spent on reading and answering questions did not correlate well with the learning factors, which implies that simply using the absolute amount of time spent on task may not be a good factor to use in predicting a student's learning outcomes, and perhaps taking into account how the time was used by students would be a better factor to consider. Furthermore, most of the implicit indicators we recorded had significant correlations with one or more of the direct motivation survey questions in Figure 1 that we asked after each reading, which implies that implicit indicators may be effective in predicting the motivational state of the student using an intelligent tutor on an activity-by-activity basis.

5   Conclusion

Understanding and modeling motivation and student behavior is an important issue for intelligent tutoring systems. We proposed that some student interactions recorded by tutors can act as implicit indicators of motivation, and that these implicit indicators, as well as explicit ones like self-assessment motivation questions, can be used to predict student learning. We tested our hypothesis with a classroom study using a vocabular tutor that integrates implicit and explicit indicators. The results

show that some user actions, such as dictionary look ups and listening to words, correlate well with motivation questions and student performance. We also found that the use of *Direct* questions, specifically tailored to the tasks, rather than *General* and all-encompassing questions, was more effective in predicting student performance.

References

1. Williams, M., Burden, R. L.: Psychology for language teachers: A social constructivist approach. Cambridge University Press, Cambridge,(1997)
2. Pintrich, P.R., Smith, D.A.R., Garcia, T., McKeachie, W.: A manual for the use of the motivated strategies for learning questionnaire (MSLQ). Report, Ann Arbor (1991)
3. de Vincente, A., Pain, H.: Motivation Diagnosis in Intelligent Tutoring Systems. In: 4th International Conference on Intelligent Tutoring Systems (1998)
4. Baker, R.S.: Modeling and understanding students' off-task behavior in intelligent tutoring systems. In: 25th ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 1059–1068. ACM, New York (2007)
5. Cetintas, S., Si, L., Xin, Y.P., Hord, C.: Automatic Detection of Off-Task Behaviors in Intelligent Tutoring Systems with Machine Learning Techniques. IEE Trans. Learn. Tech. 3, 228-236 (2010)
6. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting Student Misuse of Intelligent Tutoring Systems. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 54–76. Springer, Heidelberg (2004)
7. Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi M.: Classroom success of an Intelligent Tutoring System for lexical practice and reading comprehension. In: 9th International Conference on Spoken Language (2006)
8. Dela Rosa, K., Parent, G., Eskenazi, M.: Multimodal learning of words: A study on the use of speech synthesis to reinforce written text in L2 language learning. In: 2010 ISCA Workshop on Speech and Language Technology in Education (2010)
9. Kulkarni, A., Heilman, M., Eskenazi, M., and Callan, J.: Word Sense Disambiguation for Vocabulary Learning. In: 9th International Conference on Intelligent Tutoring Systems (2008)
10. Heilman, M., Juffs, A., Eskenazi, M.: Choosing Reading Passages for Vocabulary Learning by Topic to Increase Intrinsic Motivation. In: 13th International Conference on Artificial Intelligence in Education (2007)
11. Pintrich, P.R., De Groot, E.V.: Motivational and Self-Regulated Learning Components of Classroom Academic Performance. J. of Ed. Psych. 82, 33–40 (1990)
12. Bandura, A.: Self-Efficacy. In: Ramachaudran, V.S. (ed.) Encyclopedia of Human Behavior, vol 4, pp. 71–81. Academic Press, New York (1994)