Title: Investigating Willingness to Communicate: Comparing Survey Data with a Participation Tracking System

Presented at the 57th JACET International Convention (Sendai, 2018) Program

Andrew Komasinski, PhD, Hokkaido University of Education
Stephanie Komashin, MDiv, Hokkaido University of Education

Abstract
This project evaluated the meaningfulness of a pre-class survey's ability to predict students' Willingness to Communicate (WTC) in-class as measured by their participation. The study took place in 15 core curriculum English classes taught by two instructors at a national university. At the start of each semester, students filled out a survey including questions on about their feelings towards the class they will take using a Likert scale to measure: Excitement, Worry, Confidence, Usefulness, Skillfulness, and perceived Difficulty. These features give students a way to express WTC. Throughout the semester, student participation events were recorded during each class session using an iPad and rated on an individual level as "Excellent," "Good," "Okay," "Poor," and "Sleep." These were reentered using a points system with "Excellent," "Good," and "Okay" receiving positive values and "Poor" and "Sleep" receiving negative values. Normalizing these participation activities for each class, these values were compared with the students' impressions about the class at the beginning of the class. At the 95% confidence level, this study found no correlations between student's stated perceptions of worry and class difficulty. For Instructor A, both particular classes and the aggregate indicated a correlation between students' reported excitement towards a class and its perceived usefulness and their subsequent participation. For Instructor B, one class correlated positively based on how skilled they believed they were and another class correlated negatively with student's perceived confidence, but these correlations disappeared in aggregate. From this, the study concludes that students' self-reported beliefs about WTC do not meaningfully track their subsequent willingness and ability to participate well in classes. Consequently, it suggests that alternative methods such as comparing participation in a sample class session are necessary to meaningful measure WTC.

Introduction

Students in the same class learn and master material to different degrees. Several decades of language pedagogy maintain that one major deciding factor in language achievement is student motivation. Several studies have sought to define this connection better in order to harvest the benefits of motivation to produce improved learning achievement (Gardner & Tremblay, 1994, p. 363). This motivation relates in both instrumental and integrative ways to an individual's broader life and goals (Gardner & Tremblay, 1994, p. 361).

The only way to figure out whether motivational strategies work or result in changed classroom behavior is to test strategies in the classroom (Gardner & Tremblay, 1994, p. 362; Guilloteaux & Dörnyei, 2008, p. 56). Motivation orientation of language teaching (MOLT) provides a schema to look at how motivational the arrangement of the classroom is (Guilloteaux & Dörnyei, 2008, p. 57). While MOLT directs most of its analysis of motivation to how teacher's behavior can create a positive motivational environment, it also looks for signs of highly motivated students by evaluating their *Attention*, *Participation*, and *Volunteering for teacher-fronted Activity* (Guilloteaux & Dörnyei, 2008, p. 62).

As Guilloteaux and Dörnyei (2008, p. 58) note, "the standard data gathering technique of L2 motivation research—namely, the administration of questionnaires—would not be sufficient to assess" student in-class learning behaviors. In their study, they opted against a longitudinal approach and looked at 40 different classes with 1300 total students, once per class (Guilloteaux

& Dörnyei, 2008, pp. 59–60). They did so in conjunction with surveys of students' motivational attitudes.

Another major idea in the language education literature on motivation is willingness to communicate (WTC). This research largely builds on Gardner's socioeducational framework and integrates this with larger personality traits and motivation. Yashima, building on this research, looked at as "a way to account for individual differences in L2 communication" (Yashima, 2002, p. 54). From this work, Yashima and others point to the proof of a negative relationship between student anxiety and student WTC (Yashima, 2002, p. 55). (MacIntyre & Charos, 1996). One limitation of many of these studies is the degree to which they depend on self-reporting such as that used by MacIntyre and Charos (1996, p. 13) and Yashima (2002, p. 60).

Methodology

This study builds on the motivation research by looking at students' self-reported attitudes towards a class in a pre-class survey and comparing this with students' actual participation during every of a class over a 15-week semester for fifteen different classes. In the process, this study asks two research questions. First, can researchers put any stock in student surveys about motivation? Second, were the questions asked useful for self-reporting motivation towards a class?

To overcome the limitation posed by using questionnaires to assess student in-class learning behavior, this study uses an ICT-based evaluation tool that enables an instructor to evaluate student participation while the class is happening. This tool records individual instances of participation and enables instructors to grade them and thus assesses learners' motivated behavior. This study then compares this data with students' self-assessed interest level in these courses at the beginning of the semester. This paper compares students' self-indicated WTC with their actual in-class participation in fifteen different general education English courses taught at one national university in Japan. This was accomplished through a combination of data collection and data scoring.

Data Collection

This study provides longitudinal data from fifteen different semester-length courses taken by 571 students who provided consent to research use. These courses were taught by two instructors over a period of three academic years (six semesters). The sample sources were general English classes at one national university. Students in the classes were distributed randomly with the exception of one class A-Fall-2017-2 which was composed exclusively of English majors.

The data collection took place in two phases. The first phase happened on the first day of class. Students were required to fill out a pre-class survey that included biographical information, student career and life goals, language scores, expectations about the class, and research consent. See Appendix A. This study uses the class expectations questions which are Likert-scaled questions on the student's self-assessed:

1. Excitement level ("Not at all, A little, So-so, Excited, Very")
2. Worry level ("Not at all, A little, So-so, Worried, Very")
3. Confidence level ("Not at all, A little, So-so, Confident, Very")
4. Skill level ("Zero, Little, So-so, Skilled, Very")
5. Perceived usefulness of class ("Not at all, Unuseful, So-so, Useful, Very")
6. Perceived difficulty of the class ("Easy, Normal, Hard")

The student responses were individually coded and associated with the students using a database.

Students who did not express positive consent for inclusion in research were excluded. Consent was determined by having each student read an explanation in English and Japanese explaining their rights and asking their permission (See Appendix A). In order to qualify as consenting, students need to answer both that they understand what it would mean for the

information to be used in research and that they give permission for it to be used in research. Stated another way, students who stated that they do not understand and students who stated that they do not consent were removed from the data sets.

The second phase occurred throughout the semester. An ICT-based tool supplied an in-class participation monitoring system. Throughout the class, the instructor tracks individual participation and scores it using an iPad and tap-based interface. Individual student participation can be coded as Excellent, Good, Okay, Poor, or Sleep. These grades primarily reflect individual student's WTC rather than their performance on the assigned task. In addition, student attendance can also be tracked using the same system. This system tracked the participation of each student over the course of the semester segmenting it by participation type and day.

Thus, the raw data is student's indicated WTC based on the initial survey question answers and students' actual WTC as indicated by their participation data. Students who attended only a few class sessions (i.e. dropping the course mid-semester or other causes) were excluded from the analysis to compare apples to apples.

Data Scoring

To enable comparisons and statistical analysis across multiple classes, this study transformed the data retrieved from the two tools. First, the Likert-style answers were converted to value. The formula for these transformations is as follows:

Excitement	**Worry**	**Confidence**	**Skill**	**Usefulness**	**Difficulty**	
Very	Very	Very	Very	Very	Easy	3
Excited	Worried	Confident	Skilled	Useful		2
So-so	So-so	So-so	So-so	So-so	Normal	1
A Little	A Little	A Little	Little	Unuseful		0
Not at all	Not at all	Not at all	Zero	Not at all	Hard	-1

Stated another way, if a student says they are very excited, the system will score this as a three for excitement. Similarly, a student who says they are worried will receive a numerical two for worry. The difficulty scale was distributed so that "Easy" is three, "Normal" is one, and "Hard" is minus one.

This study also translated and normalized the raw participation data for each class. The translation for participation data interpreted the different participation grades as follows:

Grade	**Score**
Excellent	2.5
Good	2
Okay	0.5
Poor	-1
Sleep	-2.5

This scoring for participation was selected to mirror actual system use. Both "Good" and "Excellent" reflect highly participatory behavior, so the scores were made close. This scoring approximates WTC. Students who volunteer to answer or who answer eagerly when called upon earn high scores whereas students who refuse to answer when called upon or take an exceptionally long time to answer simple questions received low marks. Conversely, students

who were sleeping during class, by definition, demonstrated low WTC. scores correlate roughly to the degree to which a grade indicated WTC in class rather than accuracy of answers.

The study normalized raw participation data to create a ranked distribution of participation within each class. It made sense to normalize the data for two reasons. First, the evaluation of participation in each class depends on the instructor and the mood of the class as a whole. Normalization makes it so that the participation values for any particular class can be analyzed as a distribution such that individuals within a class are distributed based on their WTC relative to their peers.

Second, normalization produces a relative scale that enables comparisons between disparate classes. Each class, even classes using the same syllabus, have minute and more significant differences in the assignments and use of class time in each class. This means that some classes had more objective opportunities for students to communicate and participate than others. Normalization makes it so that student participation levels can be understood as distance from the mean of the class in question.

Results

This study looked at the covariance of instructor-graded in-class participation and the answers to the Likert-scale style questions described in the methodology. The following table summarizes the data for the aggregate whole and for each instructor:

Instructor	Aggregate	Instructor S	Instructor A
Excitement	**0.095**	0.06	**0.1500**
Worry	0.02	0.02	0.01
Confidence	(0.03)	-0.05	-0.01
Skill	0.066	0.09	0.03
Usefulness	**0.13**	0.1038	**0.1721**
Difficulty	0.04	0.05	0.01
Number of Students	571	353	218
Critical Value at 95% Confidence Interval	0.0689	0.1044	0.1329

For the aggregate, only student's self-reported Excitement about a class correlates with actual performance in the class, but it only explains 0.9% percent of the variance. Usefulness correlates with about 1%. For all other questions, there was no meaningful correlation at the 95% confidence interval. At the 90% confidence interval, there was a slight correlation between student's self-reported claims about skillfulness vis-à-vis course content and their participation in class, but this would only explain 0.4% of the variance. The results for Instructor S were similarly beneath the threshold of significance.

Instructor A's results, however, found two significant correlations: Excitement at 0.15 explaining 2.25% of student participation and perceived Usefulness of the class explaining 2.96% of the correlation.

Looking at the data for each individual course, several classes showed correlation between WTC and different survey parameters. Distributed by class and instructor, the results were as follows:

	Excitement	**Worry**	**Confidence**	**Skill**	**Usefulness**	**Difficulty**	**#**	**Critical Value**
S-Spring 2015	0.00	0.04	**(0.35)**	0.18	(0.05)	0.02	40	0.312
A-Spring 2015	(0.02)	(0.04)	0.19	(0.03)	0.12	0.17	37	0.3246

A-Fall 2015	**0.47**	0.04	(0.18)	0.13	0.22	(0.24)	37	0.3246
S-Fall 2015	0.18	0.06	(0.09)	0.10	(0.14)	0.05	37	0.3246
A-Spring 2016	0.07	(0.03)	0.11	0.25	0.15	0.01	39	0.316
S-Spring 2016	(0.02)	(0.25)	(0.16)	0.03	0.22	0.10	37	0.3246
A-Fall 2016	0.03	0.07	(0.23)	(0.03)	(0.04)	0.05	39	0.316
S-Fall 2016-1	0.12	0.11	0.05	**0.34**	0.16	(0.04)	43	0.3008
S-Fall 2016-2	0.02	0.19	(0.07)	(0.03)	0.15	(0.05)	41	0.3081
S-Spring 2017	0.07	(0.05)	(0.04)	0.23	(0.00)	0.09	36	0.3291
S-Spring 2017-2	(0.18)	(0.01)	0.22	0.09	0.299	0.20	37	0.3246
A-Fall 2017-1	0.03	0.00	(0.03)	0.06	**0.38**	0.15	41	0.3081
A-Fall 2017-2	**0.43**	(0.02)	0.18	(0.30)	0.20	(0.10)	25	0.3961
S-Fall 2017-1	0.22	(0.03)	(0.03)	(0.10)	0.14	(0.01)	41	0.3081
S-Fall 2017-2	0.10	0.12	0.05	(0.05)	0.15	0.11	41	0.3081

The chart displays the Pearson-correlation coefficient for each survey item compared with in-class participation split by class. Five Pearson coefficients exceed the critical value for their class sizes:

1. S-Spring 2015 (Confidence about the class at -0.35)
2. A-Fall 2015 (Excitement towards the class 0.47)
3. S-Fall 2016-1 (Skillfulness towards the class 0.34)
4. A-Fall 2017-1 (Perceived usefulness of the class 0.38)
5. A-Fall 2017-2 (Excitement towards the class 0.43)

Interpretation and Conclusion

The aggregate data show that at the 95% confidence level, there are two statistically significant connection between a student's pre-class surveys and the same student's actually in class participation, and the amount of variance these explain is minute. For the generalized student, answers about whether they are worried about a class, or believe the class to be at hard or easy, have no significant relation to the degree to which they will participate in the class. Instructor A's results found a more significant correlation, but the factors of Excitement and Usefulness can be used to explain between 2% and 3% respectively. Significant correlations were found in individual classes for both instructors, but that these disappeared in aggregate for Instructor S suggest they should be seen as insignificant.

This outcome allows for three interpretations. One interpretation has negative consequences for language learning motivation research: Surveying students about their WTC provides no generalizable information about their actual WTC. This interpretation raises questions about the feasibility of researching WTC using standard instruments and the usefulness of MOLT. Some of these concerns were identified in considering the prior literature.

Anecdotally, student motivation seems highly important for student production and student learning outcomes.

To reject this interpretation, the main alternative is to find flaws in this study's methodology. Errors could appear either theoretically or practically. On the theory side, this study assumes that the student survey is informative about a student's WTC by asking about their levels of Excitement, Worry, Skill, perceived Usefulness, and perceived Difficulty for a class. Further, the study assumes that participation in class correctly identifies WTC for the purposes of the class. As argued in the theory section above, these assumptions are well-grounded.

A third possibility suggests that the basic methodology is valid, but other factors are intervening and needs to be filtered for the correspondence between student's self-reported WTC and student's actual WTC in the classroom. One major factor this study did not evaluate is the instructor's MOLT-oriented behaviors which are central to MOLT's understanding of creating a motivational learning environment (Guilloteaux & Dörnyei, 2008). Returning to the instructor-level data, this may express differences in the classroom environments for Instructor S and Instructor A.

In the case of Instructor S, two data points showed significance at the 95% confidence level for particular classes. First, S-Spring 2015 had a statistically significant negative correlation between students' reported confidence about the class on the first day, and their WTC through the semester. Second, S-Fall 2016-1 had a positive correspondence between students' reported skillfulness and their WTC throughout the semester.

For Instructor S, the two data points that showed statistical significance provide interesting material for interpretation. The negative correlation between confidence and participation in S-Spring 2015 also disappears in the aggregate. S-Fall 2016-1 and S-Fall 2016-2 were the same content taught in the same semester by the same instructor at different time slots, but only S-Fall 2016-1 had a significant positive correlation, and this disappears in the aggregate. The main variable then was the environment that the students created as a group. This suggests that course environment coming from peers has a much larger impact than student's initial feelings about a class.

Instructor A had significant correlation at the 95% confidence level was found for both Excitement and perceived Usefulness. A-Fall 2015 and A-Fall 2017-2 both had positive significant correlation between student excitement about the class and student WTC throughout the semester. A-Fall-2017-1 found a correspondence between perceived Usefulness and participation. Looked at across Instructor A's entire data set, this significance remained on both variables. This suggests that Instructor A's teacher-driven classroom activities harness students' preconceptions about their excitement and beliefs about the usefulness of their classes, but this is to rather limited percent (2-3% of performance).

The ultimate conclusion of this study is to accept the null-hypothesis: There is no correspondence between students' self-reported WTC at the beginning of the semester and student's actual in-class participation over the course of a semester.

Limitations

This study had several limitations. The largest limitation is that the students' self-reported WTC was done only once at the beginning of the semester. A more thorough study would need to have students self-evaluate their WTC each week and then compare how they participated weekly compared to their self-assessment, how their self-assessment changed over time, and how their actual participation changed over time. At the same time, giving students these chances for reflection would also transform the experiment. One compromise would be to perform the survey several times during the semester but not every week.

A second limitation is that this study assumes the instrument accurately captures WTC through class participation. It is possible that the instructors use the tool in a way that does not accurately capture participation.

A third limitation is that both instructors teach at the same university and as reputations form through word of mouth students bring social norms about how students behave in certain classes. One version of this is that students may behave strategically in classes – being active in some classes but being affected greatly by peers in other classes.

References

Gardner, R. C., & Tremblay, P. F. (1994). On Motivation, Research Agendas, and Theoretical Frameworks. *The Modern Language Journal*, *78*(3), 359–368. https://doi.org/10.2307/330113

Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating Language Learners: A Classroom-Oriented Investigation of the Effects of Motivational Strategies on Student Motivation. *Tesol Quarterly*, *42*(1), 55–77.

MacIntyre, P. D., & Charos, C. (1996). Personality, Attitudes, and Affect as Predictors of Second Language Communication. *Journal of Language and Social Psychology*, *15*(1), 3–26. https://doi.org/10.1177/0261927X960151001

Yashima, T. (2002). Willingness to Communicate in a Second Language: The Japanese EFL Context. *The Modern Language Journal*, *86*(1), 54–66. https://doi.org/10.1111/1540-4781.00136