Psychometric Adaptation of a Japanese Version of the Feelings Towards Group Work Questionnaire for Use in the Japanese SLA Context

Larry J. Xethakis

Abstract

This measurement study reports on the adaptation of the Feelings Towards Group Work questionnaire (FTGW; Cantwell & Andrews, 2002) into the Japanese population and second language acquisition domain. The original English version of the instrument was translated into Japanese, and this was then back-translated into English and compared to the original to ensure that the language used was equivalent. A data set (N = 307) was collected from university students at two universities in western Japan. Normality of test items was examined, and reliability estimates (Cronbach's alpha) of the three subscales that make up the instrument were calculated. Confirmatory factor analysis (CFA) was conducted to test the fit of the three-factor model hypothesized by the authors of the instrument, and the evidence was negative. Subsequent, and diagnostic, CFAs of the separate subscales indicated that the model fit for the scores on two of the subscales (Preference for Individual Learning and Preference for Group Learning) was unsatisfactory while the fit on the remaining subscale (Discomfort in Group Learning) was exemplary. The implications of these results for this adapted version of the FTGW questionnaire are discussed.

Introduction

The use of pair-work and group activities has been on the increase in the English as a second language classroom (Fushino, 2010) and has become an accepted part of most teacher's practice. This is not surprising considering the benefits that have been reported in the literature for group-based learning approaches. By working in pairs or small groups, learners can act as mediators who help to explain new ideas or information to co-learners (Gillies, 2003). In addition to facilitating language learning, members of a group can also serve as a pool of resources, as sources of motivation and as a base of support for each other (Dornyei & Malderez, 1997). More broadly, small group learning approaches can improve the quality of students' interactions with their peers while at the same time helping them to develop their academic skills, build self-esteem and improve their overall attitudes towards learning (Johnson, Johnson & Smith 1998;

Slavin, 1996).

In Japan, another reason for the increase in the number of teachers employing group work in their classrooms may be the stress placed on the value of such pedagogical interventions by governmental organizations. Working in small groups has been emphasized and encouraged in policies published by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) which call for the "the promotion of innovative English education" (MEXT, 2003, p. 3). More specifically, the MEXT has called on teachers to "innovate various learning formats, incorporating pair work, group work and so on as appropriate" (MEXT, 2008, p. 7).

As Cummings et al., (2015) note, "placing students in small groups, however, does not mean that the benefits of groupwork will automatically be realized" (p. 988). When working in a more student-centered classroom, as in group-based learning, the teacher must cede a degree of control, and the participation of students cannot be easily predicted (Fushino, 2010). Students require a different collection of abilities, such as social and organizational skills and even a degree of creativity, when working collaboratively, rather than the individual competence often emphasized in more teacher-centered activities (Levine & Moreland, 1990). Gorvine and Smith (2015) suggest that teachers who wish to use group-based approaches, such as collaborative learning, in the classroom would gain from an understanding of how "students' preexisting attitudes may influence their engagement with the material and final performance in the course" (p. 56).

White, Lloyd, Kennedy, and Stewart (2005) make a number of observations concerning students and their readiness to participate fully in group work. Using group-based learning in the classroom assumes that students are comfortable with the idea of working in groups and that they possess the requisite skills to do this successfully. For students who may not be cognitively or psychologically prepared to undertake group work, working in groups, rather than aiding the learning process, might create more obstacles for it. Thus, White et al. suggest that it is important for practitioners and researchers to possess "an understanding of how students feel towards group work" (p. 617).

Furthermore, working in small groups is a highly contingent social situation—characterized by Leary (1983) as a situation where an individual's response depends upon the response of others — and so it presents more challenges to communication than is found in a traditional environment (Cowden, 2010). The greater ambiguity and uncertainty typifying contingent situations, such as pair or group work, can be difficult for learners to deal with (Kondo & Yang, 1994). Such feelings may be amplified for students who are using a second language in group work because of the increased cognitive and psychological demands inherent in such a situation.

When working in groups, individual differences among group members in areas such as personality and social skills come to the fore, and play a much larger role than they do in the traditionally structured classroom. There has been much research on the role of group work

and collaborative or cooperative learning in improving Japanese students' language skills, and even on student attitudes toward working in groups. However, there have only been a small number of studies in Japan (e.g., Fushino, 2010; Nozaki, 2015) that have focused on the interplay between group work and the kinds of cognitive, affective or social factors that influence student's dispositions towards working with others. The instruments used to measure students' dispositions towards group-based learning approaches in these studies, however, were only developed for, and used in, these studies themselves, and their psychometric qualities have not been established in further studies; or even, and in some cases in the initial study itself (e.g., Nozaki, 2012, 2015). For this aspect of group-work research to be more fully pursued, as well as to provide educators with a means to undertake pedagogical interventions for students who may be "at risk" when working in groups, there is a need for evidence-based measures of students' dispositions towards working in groups.

White et al. (2005) point to Cantwell and Andrews (2002) as perhaps the most relevant study in exploring the role which students' cognitive, social and psychological differences play in affecting their feelings toward group work. In this study, a number of cognitive and psychological factors were hypothesized to underlie students' dispositions towards working in groups, including their achievement goals, metacognitive awareness, need for affiliation and feelings of social anxiety. To help uncover the impact of these factors on students' dispositions towards group work, the authors developed the Feelings Towards Group Work (FTGW) self-report inventory using Exploratory Factor Analysis (EFA) on data from a pilot study.

The FTGW (appended to the end of this paper), which is the focus of this psychometric study, assesses three aspects of students' dispositions towards group work, namely, a preference for individual learning, a preference for group learning, and feelings of discomfort in group learning (Cantwell & Andrews, 2002). The instrument comprises 18 items, divided between three subscales (Preference for Individual Learning (PIL), 7 items; Preference for Group Learning (PGL), 7 items; and Feeling of Discomfort in Group Learning (DGL), 4 items) Items are responded to on a Likert-scale ranging from 1 to 5. Reliability estimates reported for the three subscales in the original study were .78, .71 and .60, respectively.

In Cantwell and Andrew's (2002) study, the target population was Australian secondary students, and years seven, nine and eleven were sampled to represent this educational range. Using the FTGW, Cantwell and Andrews found a strong positive correlation between students' metacognitive awareness and their preferences for working in a group. In addition, they found a strong negative correlation between students' feelings of social anxiety and students' preferences for working in groups. Those students with the highest levels of social anxiety expressed a feeling of discomfort when working in groups.

The FTGW has, since Cantwell and Andrew's (2002) initial study, been used for research in other populations and in a number of studies with university students. White et al., (2005)

was the first of these. The study examined the attitudes of Australian university students in two different programs (IT and Pharmacology) towards group work and group assessment using an adapted version of the FTGW, and found that students in both programs expressed a preference for group learning. The reported reliability estimates for the three subscales of the FTGW in this study were .82, .50 and .64, respectively. The next use of the FTGW was in Forrester and Tashchian's (2010) study which examined the effects of personality on attitudes toward academic group work among business students at an American university, using two of the three subscales of the FTGW (the PGL and DGL), along with the Neo Five-Factor Inventory (Neo-FFI; Costa & McCrae, 1992). Their overall findings, that psychological factors, in this case personality traits, affect dispositions towards group work, were in line with those of Cantwell and Andrews (2002). In particular, Forrester and Tashchian found that two of the five traits expressed on the Neo-FFI, namely extroversion and neuroticism, had strong positive and negative correlations, respectively, with students' preferences for working in groups, as well as their ability to function well in groups and participate fully in the learning process. In this study, the reliability estimates for the two FTGW subscales used were found to be .65 for the PGL scale and .63 for the DGL scale. Most recently, Gorvine and Smith (2015) employed the FTGW, together with instruments to measure students' attitudes toward studying statistics, to examine the performance of American university students' in a statistics course while studying under a collaborative learning approach. They found that students who scored high on the PGL subscale and who had less anxiety towards learning statistics also had better learning outcomes in this learning environment. However, no psychometric results for the FTGW were reported in this study.

In addition to the FTGW's migration into the university population, it has also been used in non-English speaking populations as well. Gasaymeh, Kreishan and Al-Dhaimat (2014) translated the instrument into Arabic, in order to examine university students' attitudes towards group learning in computer training courses, and found differences in students' preferences that were related to their course of study. For the two subscales of the FTGW used in this study, namely, the PGL and DGL, the reliability estimates were reported as .73 and .66, respectively. The FTGW has also been used for research on group work in EFL classes (in the university population) as well. Maesin, Mansor, Shafie, and Nayan (2009) employed an instrument adapted from the FTGW to examine the learning preferences of students in English classes at a university in Malaysia. The students in this study expressed a strong preference for group learning in their classes. The authors did not report whether the instrument was translated or whether the English version was used, however. Furthermore, reliability estimates for scores on their instrument were not reported.

The widest use of the FTGW outside of the English language domain has been in Greece. Goudas, Magotsiou and Hatzigeorgiadis (2009) developed a substantially modified version of the FTGW and examined the psychometric properties of its scores. The authors performed a

confirmatory factor analysis (CFA) on their version of the FTGW, re-named the G-FTGW, which showed that the scores on this instrument displayed a clear three-factor structure, similar to that proposed in Cantwell and Andrew's (2002) original study. In addition, the authors reported the values for the internal reliability of the subscales (alpha for each of the three subscales was .83, .87, and .95, respectively), test-retest reliability, as well as convergent and criterion-related validity. The G-FTGW has been employed in studies (Angeli & Tsaggari, 2016; Goudas & Magotsiou, 2009; Lemonia & Dimitris, 2017) which have examined students' dispositions toward group-based learning approaches in university, secondary and elementary student populations.

Finally, while the FTGW itself has not been used in Japan or the Japanese EFL context, Cantwell and Andrews (2002) study informed the instruments developed by Nozaki (2012, 2015) for investigating the use of group work in university as well as in primary education contexts. The first of these studies (Nozaki, 2012) looked at the role of cooperative learning approaches on Japanese university students' attitudes towards asking their peers questions in class, and found some positive effects on the behavior after students' exposure to this learning approach. In Nozaki's second study (2015), the interplay of primary students' academic year and achievement goals on their perceptions of group work were examined. (It is interesting to note that achievement goals was one of the factors that Cantwell and Andrews examined in their original study). It was found that the mastery goals of younger students disposed them more strongly towards group learning, while those with performance-avoidance goals were more disposed towards individual learning. In addition, Nozaki found that older students were more disposed towards individual work, but that their achievement goals had little effect on this preference.

Given 1) the present use of pair and group work in teachers' practice, 2) the push for the increased use of group-based approaches, and 3) the important influence of students' psychological and cognitive differences on their ability to engage in and benefit from group work, there is a need for evidence-based measurement of students' dispositions towards working in groups. Such evidence-based measurement provides a foundation for the pursuit of research with secure empirical foundations. By undertaking an examination of the psychometric properties of a Japanese version of the FTGW, the present study is an incremental contribution to this research agenda.

Methodology

The methodology for this study is reported in terms of the instrument itself and its design, the participants and data collection procedure, and the analytical procedure, which focuses on the statistical and psychometric methodology for examining the scores generated by the instrument.

Instrument

The FTGW (Cantwell & Andrews, 2002) comprises 18 items based on a five-point Likert scale, with 1 semantically anchored by the statement *not true at all of me* and 5 by the statement *very true of me* (see Appendix). The instrument is hypothesized to comprise three subscales measuring aspects of respondents' feelings towards groups work — the PIL subscale (Items 1, 5, 6, 12, 14, 16 and 18), the PGW subscale (Items 3, 7, 8, 9, 11, 13 and 15) and the DGW subscale (Items 2, 4, 10 and 17). The latter subscale (four items) is shorter than the first two subscales (seven items). Scores on each of the three subscales are calculated separately and are not summed for a composite score on the instrument as a whole.

As the instrument was originally developed for use in an English speaking context, the initial step in adapting the FTGW into the Japanese SLA context was the forward- and back-translation of the instrument following the guidelines issued by the International Test Commision (Hambleton et al., 2005). The forward translation was done by the author working in collaboration with a native Japanese speaker, while the backtranslation was done by two native English speakers fluent in Japanese. All of the translators were professors or instructors at local universities, with some experience in testing. The back-translated versions were compared to the original English version and a few inconsistencies were identified. For example, with respect to Item 8, "*It is important that other group members take responsibility for my learning as well*," both back-translators translated the Japanese version to read that the respondent had a responsibility to the other group members, rather than the reverse as is stated in the original item. The points of inconsistency between the translated and back-translated versions were modified and all the translators came to agreement on the content of the items. It was then decided that the Japanese version of the instrument was ready for psychometric evaluation in the Japanese SLA context.

Participants and Data Collection Procedure

There were a total of 363 participants in this study. From the total number of responses, 32 records were found to have missing data and were subsequently removed from the data set. As determined by inspection, there was no discernible systematic pattern to the missing responses and therefore removing these records did not affect the properties of the data set in a systematic manner. A further 24 records were removed due to an obvious pattern response, such as all 1s being checked. The analysis described below was based on the remaining 307 records. The participants came from two universities in western Japan, studying in the fields of civil and environmental engineering (n = 62), computer science and electrical engineering (n = 128), applied chemistry and biochemistry (n = 21), medical care and welfare engineering (n = 19), plant science (n = 8), animal science (n = 11), bioscience (n = 7), business administration (n = 20), tourism management (n = 13), and law (n = 18). The age of the participants ranged from 17 to 25, with 53.7% of the respondents in their 3rd year of study. There were 70 female and 236 male

respondents (1 respondent did not indicate his/her gender).

In addition to the above information concerning gender, age, university and faculty, respondents were asked to indicate whether or not they had experience with group work in junior high school, high school and university English classes, and also the frequency of their exposure to group work (never, hardly ever, occasionally, sometimes, often, frequently, or almost always).

The respondents' participation in the survey was completely voluntary and had no effect on the students' assessment. A form asking for the participants' informed consent was included at the beginning of each questionnaire and it was clearly stated in Japanese that those students not consenting to take part could do so simply by not completing the questionnaire. The time required for the students to complete the questionnaire was approximately 10 minutes, but no time limit was stipulated as part of the administration of the instrument.

Analytical Procedure

The data collected from the respondents was entered into a Microsoft Access 2016 database. IBM/Statistical Package for the Social Sciences (SPSS) software (Version 21) was used to calculate descriptive statistics and reliability estimates (Cronbach's alphas) for the scores. The confirmatory factor analysis (CFA) was conducted using AMOS (Version 21). The data was initially considered from the point of view of descriptive statistics, with a focus on univariate normality (i.e. skew and kurtosis). After this, reliability estimates were calculated prior to carrying out a CFA on the instrument as a whole, as well as on each of the individual subscales taken independently.

In order to evaluate skew and kurtosis, the critical ratio was determined by dividing the skew and kurtosis values by the respective standard errors for each of the 18 items. In evaluating the skew and kurtosis, the results of the calculation were compared with a minimum interpretive criterion of 3.0, as well as a stricter criterion of 2.0, both stipulated in advance by the author.

With respect to calculating reliability estimates (Cronbach's alphas), confidence intervals (95%) were also calculated following recommendations by Fan and Thompson (2001). A value of greater than .70 for scale reliability, based on Nunnally and Bernstein's (1994) criterion, was adopted for interpretation of alpha. It is important to note, however, that these calculated values for alpha are not the central part of the analytical procedure, and this role is reserved the CFA. This is because CFA is more powerful than alpha as a diagnostic method for the structural validity of the scores generated by an instrument. Alpha cannot be used to establish the unidimensionality of subscales (Gerbing & Anderson, 1988), whereas a CFA employing a common factor model (with items hypothesized to load on only one factor, and the error between indicators stipulated to be uncorrelated) can be used to determine this.

Results

The results of the analysis of the scores generated by the FTGW instrument are presented in three parts. In the first part, the descriptive statistics for the scores are presented, which summarize the properties of the sample and the distribution of scores on each item. These include the means (as indicators of central tendency), the standard deviation, and skew and kurtosis for each item. In the second part, the reliability estimates (Cronbach's alpha and associated confidence levels for alpha) for each of the subscales are presented. The third part covers the results of the CFAs, which include the instrument as a whole (correlated model), as well as on each of the individual subscales taken independently (uncorrelated model as there is only one factor in each case).

Descriptive Statistics

Table 1 below presents the descriptive statistics with respect to score distribution. The highest mean was for Item 15 (3.96) and the lowest mean was for Item 14 (2.40). Standard deviation ranged from 0.739 (Item 13) to 1.097 (Item 02).

Table 1. Item Means, Standard Deviation, Skew and Kurtosis for each Item

Item	Mean	SD	Skewness		Kurtosis	
			Statistic	Std. Error	Statistic	Std. Error
01	2.63	1.012	0.236	0.139	-0.437	0.277
02	3.41	1.097	-0.502	0.139	-0.403	0.277
03	3.42	0.909	-0.290	0.139	-0.097	0.277
04	2.52	0.818	0.556	0.139	0.339	0.277
05	2.79	0.960	0.174	0.139	-0.415	0.277
06	2.88	0.933	0.000	0.139	-0.255	0.277
07	3.44	0.921	-0.418	0.139	-0.066	0.277
08	3.16	0.889	-0.382	0.139	0.173	0.277
09	2.63	0.858	0.350	0.139	0.417	0.277
10	2.67	0.935	0.257	0.139	-0.230	0.277
11	3.37	0.996	-0.140	0.139	-0.284	0.277
12	2.62	0.991	0.345	0.139	-0.248	0.277
13	3.64	0.739	-0.474	0.139	0.547	0.277
14	2.40	0.839	0.344	0.139	0.047	0.277
15	3.96	0.783	-0.672	0.139	0.830	0.277
16	2.60	0.885	0.042	0.139	-0.117	0.277
17	2.93	0.969	-0.079	0.139	-0.452	0.277
18	2.47	0.930	0.150	0.139	-0.403	0.277

Table 2 shows the results for the calculations of the critical ratio for skew and kurtosis for each item. This calculation was done by dividing the respective values for skew and kurtosis for each item by their respective standard errors. The resultant values were then compared to the two criteria (2.0 and 3.0) mentioned above in the Methodology section. Those values which failed to meet the more relaxed standard of 3.0 are marked with two asterisks, while those which failed to the stricter standard of 2.0 are marked with one asterisk.

Table 2. Critical Ratios for Skew and Kurtosis for each Item (Absolute Value)

Item	Skewness	Kurtosis
	Calculated Value	Calculated Value
01	1.70	1.58
02	**3.61	1.45
03	*2.09	0.35
04	**4.00	1.22
05	1.25	1.50
06	0.00	0.92
07	**3.01	0.24
08	*2.75	0.62
09	*2.52	1.51
10	1.85	0.83
11	1.01	1.03
12	*2.48	0.90
13	**3.41	1.97
14	*2.47	0.17
15	**4.83	**3.00
16	0.30	0.42
17	0.57	1.63
18	1.08	1.45

Note: *Test item is skewed at the 2.0 threshold. **Test item is skewed at the 3.0 threshold.

As shown in Table 2, in terms of skew, 8 items (44.4%) fell below the 2.0 threshold, 5 items (27.8%) fell below the more relaxed 3.0 threshold, while 5 items (27.8%) failed to meet even the more relaxed threshold. In terms of kurtosis, the items performed better, 17 of the items (94.4%) met the strict 2.0 threshold, while only one was on the threshold of the more relaxed value. The normality of the scores for the items making up the FTGW was not reported in the original study (Cantwell and Andrews, 2002), and thus, it is difficult to determine if the non-normality of some of the items occurred as a result of the process of adapting the instrument to the Japanese SLA context or if a similar degree of non-normality was found in the original version.

Reliability Estimates

The reliability estimates (Cronbach's alpha) for scores on each of the three subscales proposed by Cantwell and Andrews (2002) for the FTGW are presented in Table 3.

Table 3. Reliability Estimates, Confidence Intervals for Alpha (95%), Scale Means, and Scale Standard Deviations for Scores on the Subscales of the FTGW

Subscale	Cronbach's alpha	95% Confidence Intervals for Cronbach's alpha		Scale Mean	SD for Scale
		Lower Bound	Upper Bound		
PIL	.794	.757	.827	18.38	4.387
PGL	.565	.487	.636	23.62	3.222
DGL	.519	.425	.601	11.53	2.456

Table 4 below presents the values for Cronbach's alpha for each of the subscales on the FTGW as reported in the literature as well as the results for this study for the purpose of comparison. For the first of the three subscales, the PIL subscale, the lower bound of the 95% confidence interval for alpha was above the .70 threshold (discussed in the methodology section above), and was also the highest of those reported in this study. As can be seen in Table 4, this follows a trend found in the literature, with this subscale also having the highest alpha of the three subscales in both Cantwell and Andrews (2002) and White et al., (2005). The value for the second subscale, PGL, fell far below this threshold, however, with even the upper bound of the 95% confidence interval below this value. This differs from the trend found in the literature, where the alpha value for the PGL subscale tends to hover at or below the threshold value in most of the reported studies, with the exception of White et al., (2005), as shown in Table 4. The value for the remaining subscale, DGL, fell below the threshold as well. It should also be noted that the alpha value for this subscale in the original instrument, reported by Cantwell and Andrews' (2002) as .60, also fell below the threshold value and was the lowest value of three subscales in the original instrument as well. The reported values in the literature tend to follow this trend, with this subscale having the lowest value in all of the studies, with again, the exception of White et al. (2005).

Table 4. Comparison of Cronbach's Alpha for FTGW Subscales Reported in the Literature

Subscale	Cronbach's alpha				
	This study	Cantwell & Andrews	White et al.	Forrester & Tashchian	Gasaymeh et al.
PIL	.79	.78	.82	–	–
PGL	.56	.71	.50	.65	.73
DGL	.52	.60	.64	.63	.66

The low values of the PGL and DGL subscales would tend to bring the reliability of the subscales into question. However, in the case of the DGL subscale, which comprises only four items, the small number of items may be a factor. As has demonstrated by Cortina (1993) and Green, Lissitz and Muliak (1977), alpha tends to have a bias towards larger numbers of items on a subscale, and thus subscales with fewer items tend to produce lower alphas. This feature of alpha may provide a justification for a relaxation of Nunnally and Bernstein's (1994) .70 benchmark when interpreting the value for alpha in the case of this subscale. However, this line of argument cannot be employed when interpreting the alpha value for the PGL subscale of the instrument and so the reliability of this subscale on this adapted version could still be called into question.

Confirmatory Factor Analysis

In order to test the hypothesized three-factor structure of the FTGW directly, a CFA was performed. In addition, separate CFAs were carried out on each of the individual subscales taken as single-factor models. In evaluating the fit of the models a combination of the chi-square and a number of fit indices were employed. The reason indices are used in conjunction with chi-square is to help overcome the chi-squares tendency to over-reject models due to its sensitivity to sample size. Hu and Bentler (1999) have recommended empirically derived cut-off values for a number of indices, to be used in conjunction with each other, and these have been adopted in this study.

Four indices were employed for each of the models examined in this study: the root mean square error of approximation (RMSEA), the standardized root mean squared residual (SRMSR), the Tucker-Lewis index (TLI), and the comparative fit index (CFI). These four indices are among the most commonly used in applied research (Brown, 2015). The RMSEA is a parsimony correction index and thus it rewards simpler models over more complex ones. For the RMSEA index, the smaller the value (that is closer to zero), the better the fit. The SRMSR is an absolute fit index and takes the average of the standardized residuals, or the differences between the correlations as predicted by the model and those actually observed in the data (Brown, 2015). For the SRMSR index as well, the smaller the value, the better the fit. The TLI and CFI are comparative fit indices that evaluate the fit of the stipulated model in relation to the baseline, or independence, model. The CFI has a range of 0 to 1, with a higher value indicating better fit. Brown (2015) notes that the TLI is non-normed and therefore its value can fall outside of the 0 to 1 range, with a higher value indicating better fit. Table 5 outlines the calculated values for each of the indices employed in the analysis of the goodness-of-fit for each of the models in this study.

Table 5. Comparison of Goodness-of-fit Indicators for FTGW Three-Factor Model and Subscale One-Factor Models

Index		Scale / Subscale			
	Cut-off value	FTGW	PIL	PGL	DGL
TLI	(>.95)	.729	.875	.691	1.034
CFI	(>.95)	.767	.917	.794	1.000
RMSEA	(<.06)	.081	.105	.089	.000
SRMSR	(<.08)	.0864	.0594	.0624	0163
Mardia's coefficient	(< 5)	29.82	15.29	13.06	2.188
Chi-square	(value)	399.83	60.99	47.62	1.20
(probability level)		.000	.000	.000	.548

Three-Factor Model

This is the model proposed by Cantwell and Andrews (2002) as a result of EFA. The model comprises three factors purported to measure aspects of respondents' dispositions towards groups work, the PIL subscale (Items 1, 5, 6, 12, 14, 16 and 18), the PGL subscale (Items 3, 7, 8, 9, 11, 13 and 15) and the DGL subscale (Items 2, 4, 10 and 17). The model had 171 distinct sample moments, 39 distinct parameters to be estimated, and 132 degrees of freedom, meaning that the model was overidentified. Figure 1 shows the CFA model employed in this study to examine the fit of the originally hypothesized three-factor model.

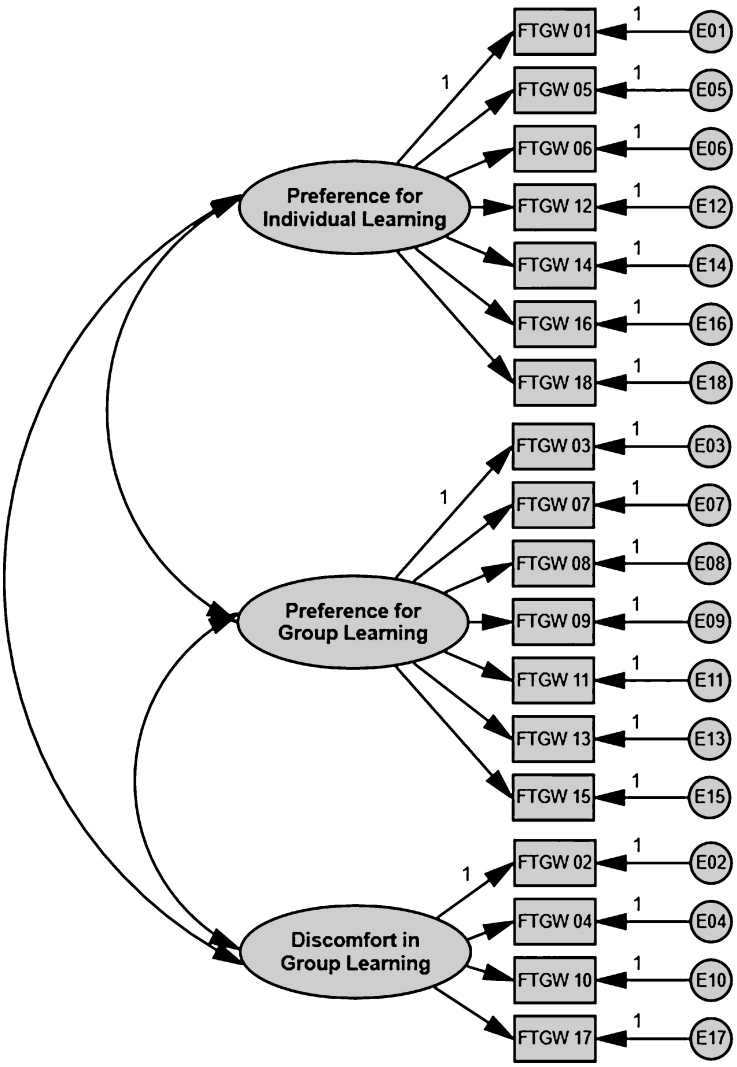

Figure 1. Three-factor model of FTGW used in CFA (based on Cantwell and Andrews' [2002] original EFA). The large ellipses are used to represent underlying constructs. The small circles are used to represent measurement error. Each item is linked to its hypothesized underlying construct by a single-arrow to show the relationship between the two. Each of the underlying constructs are linked to the others by double-headed arrows to show the hypothesized correlation of the constructs.

The results for this model were as follows, with Hu and Bentler's cut-off values given in parentheses: TLI .729 (>.95), CFI .767 (>.95); RMSEA .081 (<.06); SRMSR .0864 (<.08). The chi-square value was 399.838 with a probability level of .000. Additionally, Mardia's coefficient was used to assess the multivariate normality of the data. The critical ratio for this model was 29.82. For this measure, any value over five indicates multivariate nonnormality. These results

indicate that the data set fits the model insufficiently and therefore that the model hypothesized by the original authors is problematic in the present adaptation of the instrument.

Single-Factor PIL Model

This is model comprises only one of the subscales in the originally hypothesized model, the PIL subscale (Items 1, 5, 6, 12, 14, 16 and 18). The model had 28 distinct sample moments, 14 distinct parameters to be estimated, and 14 degrees of freedom, and thus the model was overidentified. The CFA model used in this study to examine the PIL subscale is shown in figure 2.

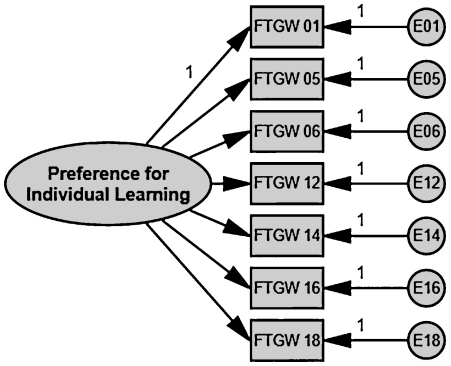

Figure 2. Single-factor PIL model used in CFA (based on one of the factors emerging in Cantwell and Andrews' (2002) original EFA).

The results for this model were as follows, with Hu and Bentler's cut-off values given in parentheses: TLI .875 (>.95), CFI .917 (>.95); RMSEA .105 (<.06); SRMSR .0594 (<.08). The chi-square value was 60.990 with a probability level of .000. Mardia's coefficient for this model was 15.29, indicating multivariate nonnormality in the data. These results indicate that the data set fits the model insufficiently and therefore that this single-factor model is problematic in the present adaptation of the instrument.

Single-Factor PGL Model

As with the above model, this model also comprised only one of the subscales in the originally hypothesized model, the PGL subscale (Item 3, 7, 8, 9, 11, 13 and 15). The model had 28 distinct sample moments, 14 distinct parameters to be estimated, and 14 degrees of freedom, and so the model met the criteria for overidentification. Figure 3 shows the CFA model use to evaluate the PGL subscale in this study.

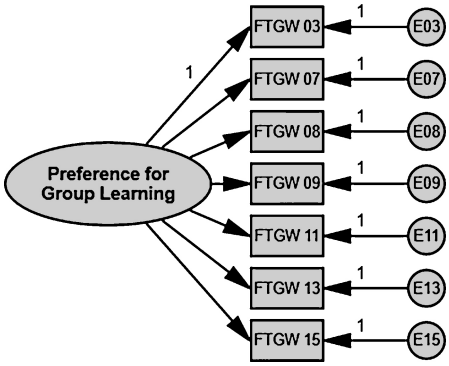

Figure 3. Single-factor PGL model used in CFA (based on one of the factors emerging in Cantwell and Andrews' (2002) original EFA).

The results from the calculations of the fit indices for this model were as follows (Hu and Bentler's cut-off in parentheses): TLI .691 (>.95), CFI .794 (>.95); RMSEA .089 (<.06); SRMSR .0624 (<.08). The value of chi-square was 47.624 with a probability level of .000. Multivariate nonnormality in the data was indicated by the value for Mardia's coefficient for this model, 13.06. As a result, the model hypothesized by the original authors for this subscale is also problematic in the present adaptation of the instrument, as the present data set lacks sufficient fit with the model.

Single-Factor DGL Model

The final model similarly comprises only one of the subscales in the originally hypothesized model, the DGL subscale (Items 2, 4, 10 and 17). The model had 10 distinct sample moments, 8 distinct parameters to be estimated, and 2 degrees of freedom, meaning that the model was overidentified. The CFA model used to evaluate the fit of the DGL subscale is shown in figure 4.

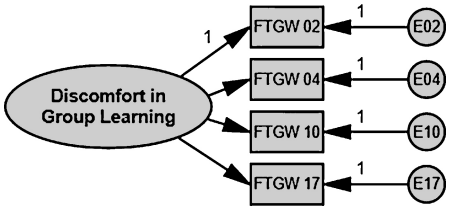

Figure 4. Single-factor DGL model used in CFA (based on one of the factors emerging in Cantwell and Andrews' (2002) original EFA).

For this model, the results of the fit indices were as follows (Hu and Bentler's cut-off values in parentheses): TLI 1.034 (>.95), CFI 1.000 (>.95); RMSEA .000 (<.06); SRMSR .0163 (<.08). In contrast to the three models examined above, the values for all four of the reported indices suggest that this model is a good fit for the data. The chi-square value was 1.204 with

a probability level of .548. This result shows that there is not a significant difference between the dimensionality of the data and that predicted by the model. Additionally, the critical ratio calculated for Mardia's coefficient for this model was 2.188, indicating multivariate normality in the data. All of these measures suggest that there is high degree of probability that this single-factor model fits the data and so could be employed in research and in practice.

Discussion

The purpose of this study was to examine the psychometric properties of scores generated by a version of the FTGW adapted for use in the Japanese EFL context. The FTGW has been used in a range of populations to investigate the cognitive, affective and social factors that influence students' dispositions towards working in groups, and the main goal of this study was to facilitate the extension of this research trajectory into the Japanese population by providing evidence, either negative or positive, for the plausibility of the measurement model in this population. If the scores from this data set were shown to confirm the structure hypothesized by Cantwell and Andrews (2002), the FTGW could become a useful tool, with an evidence-based foundation, for measurement of Japanese EFL students' dispositions towards working in groups

However, the results reported in this paper indicate that there was not a good fit between the data gathered in this study and the original, three-factor model hypothesized by Cantwell and Andrews (2002). Moreover, as a result of the diagnostic CFAs conducted on the scores for each of the subscales, it was found that two of the three subscales were a poor fit when treated as single scales. The results were not wholly negative however, as one of the three subscales, DGL, displayed meritorious fit when isolated and treated as a single scale.

While some of the results may have been less than satisfactory, the data gathered in this study nevertheless constitute evidence, and evidence, whether positive or negative, is preferred over no evidence at all. Furthermore, the evidence which is provided serves as a guide for further steps in the adaptation of the FTGW for use in the Japanese EFL context. The first step in this process would be the determination of changes that could be made to the items comprising the instrument; i.e., revising the translations, adjusting the content, or even, if called for, removing particularly troublesome items and so forth. These changes would bring about a new adapted version of the FTGW, whose structural resemblance to Cantwell and Andrews (2002) hypothesized structure could be evaluated on a new set of scores. To this end, it is useful to examine the properties of the dataset in the present study, giving consideration to the following: 1) the normality or nonnormality of score distributions in the data; 2) the reliability estimates for subscales; and 3) the interpretation to be placed on the results of the CFA carried out on the entire instrument, as well as those CFAs carried out on each of the individual subscales treated as single scales. These examinations are discussed in turn below.

Normality of the Data

In this area, there were both satisfactory and unsatisfactory results. Among the former, it is noteworthy that there was very little kurtosis shown in the scores, with only one item, Item 15, reaching the value of the more relaxed 3.0 threshold. The remainder of the items fell below the stricter 2.0 threshold and fully half of the items met, or were extremely close to a threshold value of less than or equal to 1.0, which can be considered meritorious. This means that for these items the distribution of the scores in terms of kurtosis was not distorted.

There were, however, issues with the degree of skew for some of the items. While five of the items met the criteria to be considered meritorious, and an additional three items met the 2.0 threshold, the remaining ten items exceeded this threshold. Five of these remained below the less stringent 3.0 threshold, but five items exceed even this threshold, showing a problematic degree of skew. Items 4 and 15 were particularly notable here, with values of 4.00 and 4.83, respectively. When the scores on an item display a large degree of skew, there is a loss of information which is caused by the grouping of the scores towards one or the other extreme end of the scale. In looking at the role of individual differences in influencing behavior, it is the variability in the responses that provides the information about these differences. In instances where the scores on an item are clustered so closely together, such as in Items 4 and 15, very little information is obtained. This makes it difficult to perceive differences between individuals in regards to the particular aspects of the underlying constructs that the items are attempting to represent and thus the responses on these items shed much less light on the respondents than they otherwise might.

One means to address this issue would be to revise the translations of these items with careful consideration of words or phrases that might be influencing the respondents to cluster around certain points on the scale. Another means of rectifying this issue could be to adjust the Likert scale from five points to a greater number, such as seven or possibly even higher. This adjustment would create a more finely grained scale and make it more sensitive to the respondents' dispositions, allowing the scores to possibly spread over a wider scale width and helping to overcome the issues with skewed distributions. However, this solution brings with it its own possible difficulty. It raises the question of whether respondents can actually make such a fine-grained distinction concerning their dispositions, or whether this will simply introduce more error into the data as respondents haphazardly try to make such distinctions.

Examining the content of the items with skewed distributions could provide a third process for solving this problem. As mentioned above, Item 15 and Item 4 had the greatest degree of skew. Interestingly, these two items were skewed in opposite directions, with Item 4 exhibiting a moderate degree of positive skew and Item 15 a severely negative skew. Item 15 was also the one item for which the kurtosis reached the 3.0 threshold. This means that in addition to the negative skew, the responses had a very sharp peak to them, with 165 respondents (53.7%) stating that the item was true of them (number four on the Likert scale). For this item, "*It is*

best when each person helps each other within a group," the overwhelming positive nature of the response to this item (a total of 237 respondents said the statement was either true or very true of them) may be due to a number of underlying causes. The first of these may have to do with the different target populations in Cantwell and Andrews' (2002) study and this study, i.e. students in secondary education and university students. The greater degree of maturity found in the university student population may make it seem more obvious to respondents that helping each other would be a prerequisite for success in group tasks. A second possible cause may be the greater degree of group work experience found for many of the respondents in this study. Large majorities of the respondents in this study had exposure to group work in English classes in junior high school (74.9%), high school (60.3%), or university (88.6%). These experiences might also make it appear more natural to the respondents in this study that each group member should help others within the group. The third plausible cause for this skew may be related to cultural factors. Japanese culture is well-known for its emphasis on the group over the individual, and Japanese students arguably have a greater degree of experience working together in informal groups from the time they enter the educational system, such as when cleaning their classrooms or other areas of their schools. These kinds of cultural elements might predispose students towards the strong agreement seen in the responses to this item. Finally, there is an element of this statement which presents a kind of truism, regardless of culture, and even as a kind of platitude were one to take a more critical tone.

Similar factors might play a role in explaining the distribution of responses on Item 4, "*I often find it difficult to understand what the group task is.*" A mere 34 respondents said the statement was either true or very true of them. In this case, the greater maturity of the university students in this study could imply that they are cognitively more able to deal with the complex instructions involved in group tasks, and thus the task would be more easily understood by a larger number of students. Group work experience might play an even larger role here, and it may be that English (or other foreign language) classes are peculiar in this regard. Many of the pair- or group-work tasks commonly used in English classes are both repetitive in nature (the students repeat the task a number of times) and also tend to employ similar processes and procedures to accomplish the task at hand. With many years of exposure to these types of activities, students in the Japanese EFL context may come to possess a basic understanding of the types of tasks that they have encountered. This would assist their understanding of the general outlines of such tasks when presented in new contexts, lending strength to students' sense of comprehension. Finally, it is well known, anecdotally, that Japanese students tend to ask classmates for confirmation before beginning almost any task in the English classroom, and this tendency might explain a greater degree of willingness to ask about or explain any unclear aspects of a group task. This would again provide a greater degree of understanding of the group task.

It would be hasty to suggest that items with strongly skewed distributions be removed

from their respective subscales based only on the above results, and in considering these issues and their possible solutions, it must be noted that neither Cantwell and Andrews (2002), nor any other subsequent studies employing the FTGW, have reported figures on the degree of nonnormality in their scores. Thus, these issues might not lie solely in the data for this study and the adaptation of the instrument undertaken here. That being said, a case could be made that in instances where there is strong agreement among participants, such as for Items 4 and 15, the item is contributing very little new information on the underlying construct and therefore could, or even should, be replaced with an item that produces a greater degree of variation in scores.

While acknowledging the above issues regarding the unsatisfactory degree of skew in the scores for a number of items on the FTGW, it should also be recognized that a number of the results were good or even meritorious. As mentioned above, the degree of skew on five of the items met the criteria to be considered meritorious. Item 6, "*I prefer to work within a group rather than work alone,*" is a good example with a skew of 0.00 and a kurtosis of 0.92. One reason for the meritorious distribution of scores on the item may be that that the content of the item appears to be very clear and easily understandable, presenting a simple choice between two alternatives. It may also be that in asking about a preference, the item draws on a more emotionally-based, affective response rather than a cognitively-based one. For respondents of any age, gauging the strength of a feeling of attraction or aversion could possibly be easier than gauging their own cognitive abilities or degree of social skills.

Reliability Estimates (Cronbach's alpha)

Next, this discussion will turn to an examination of the reliability estimates (Cronbach's alpha) for the three subscales of the FTGW. While alpha is the most commonly reported measure of reliability for scores produced on subscales of a psychometric instrument, the values returned for it, and their interpretation, should not take precedence over the results of more powerful techniques, such as CFA. One of the primary reasons that alpha is reported in this study is to allow for comparisons with previous studies.

Concerning the values for alpha reported in this study, only one of the three subscales, the PIL subscale, had an alpha value that was above the commonly accepted .70 threshold. This scales' alpha value was also the highest of those reported in this study. As noted above, this scale tends to have the highest value of the three subscales in those studies reporting the value. It is interesting to note that even with its consistently high alpha values, the scores on this subscale did not display particularly good when treated as single scale in the CFA conducted in this study. This result corroborates the limitations of relying solely on alpha.

The alpha for the second subscale, PGL fell far below the .70 threshold. This value is lower than that reported in most other studies. One possible reason for the poor alpha value for this scale might be the degree of non-normality found for items comprising it, with 6 of the 8 items

above the 2.0 threshold. Taking these results together with those from the CFA, which showed that the scores on this subscale had the worst fit of three subscales, it may be that this subscale would require the greatest degree of alteration in order to reach a scale with good model fit.

The alpha value for the third subscale, DGL, also fell quite far below the threshold, and was the lowest among the three subscales. One reason for the consistently low value for alpha for the subscale could be the small number of items that comprise this subscale. As mentioned above in the Results section, alpha tends to have a bias towards larger numbers of items on a subscale (Cortina, 1993; Green et al., 1977), and thus subscales with fewer items tend to produce lower alphas. If this is one of the factors influencing the alpha value on this subscale, it could be a rationale for relaxing the .70 benchmark suggested by Nunnally and Bernstein (1994). There might be another factor at work here as well, which is the make-up of the items of the scale itself. Examining the content of the four items that make up this scale (See Appendix), it is interesting to note that three of the four items, Items 2, 10 and 17 ask about affective responses towards group work, employing words and phrases such as, feel nervous, afraid and rarely feel relaxed. However, the remaining item, Item 4, is concerned more with the cognitive complexity involved in group work. The fact that these items grouped together on the same factor in Cantwell and Andrew's (2002) EFA may not be surprising, since it was also found that both social anxiety and level of metacognitive awareness played a role in students' feelings of discomfort with group work. However, it may be that the different aspects at play here are helping to lower the alpha value, which is after all an expression of the shared correlation among them.

That being said, and in contrast to this subscale's consistently low alpha value, the diagnostic CFA performed in this study showed exemplary fit for this subscale. In cases such as this, where the value of alpha is below the threshold, but the results of the CFA are good, the CFA, as the stronger analytic tool, should take precedence. The converse is also true. Where alpha values are satisfactory, but the results from a CFA are not, as with the PIL subscale, the results from the CFA should be determinative in the final interpretations of the plausibility of the scale.

Confirmatory Factor Analysis

Moving on to the results for the CFAs conducted in this study, those for the original three-factor model hypothesized by Cantwell and Andrews (2002) will be discussed first, followed by those for the three subscales treated as single-factor models, undertaken as a diagnostic for determining which of the three subscales were the most problematic.

In determining the fit of a model using CFA, the model must perform satisfactorily on all of the goodness-of-fit indices in order to be deemed acceptable (Hu & Bentler, 1999). As can be seen in Table 5 above, the original three-factor model failed to meet the cut-off values for any of the indicators that were employed in this study to measure goodness-of-fit. This is fairly

emphatic evidence of the unsatisfactory nature of this model under the present adaptation. It is for this reason that the subsequent analysis of the subscales, as individual single-factor models, was conducted in order to assist in determining which of the subscales were more problematic and which of them showed possible potential for separate use. This analysis was conducted primarily for the purpose of the further development of the FTGW in the Japanese EFL context.

The first of these diagnostic models was the PIL subscale. The seven items in this subscale (See Appendix) showed very little non-normality in their distributions, with four of the seven items falling below the 1.0 threshold and only two items above the 2.0 threshold. As mentioned above, this was also the only subscale whose alpha value surpassed the suggested .70 criterion value. However, this subscale performed poorly in the CFA. The SRMSR value was below the cut-off value and therefore acceptable, but the other three indices did not meet the criteria set by Hu and Bentler (1999). The CFI value, although not satisfying Hu and Bentler's threshold of greater than .95, did fall above .90 and Brown (2015) notes that this "may be indicative of an acceptable model" (p. 75). Nonetheless, Hu and Bentler's cut-offs, which are authoritative, were adopted for this study, and furthermore, their cut-offs are to be use in triangulation (i.e. positive outcomes on all four indices) and so a marginal outcome on the CFI by itself would not suffice to indicate model fit. In addition, the value for the TLI on this single factor model, while better than that for the entire three-factor model discussed above, was still far below the cut-off value while the value for the RMSEA was far beyond the acceptable threshold (<.06). The triangulation of all four indices provide strong evidence that, at least in this adapted version of the subscale, there was an unsatisfactory fit between the suggested model and the scores.

The second single-factor model examined was the PGL subscale. This scale had a number of issues. The scores on this particular subscale exhibited a large degree of non-normality, with six of the seven items (See Appendix) exceeding the 2.0 threshold and three of these above the 3.0 threshold. It also possessed the lowest value among the three subscales for its reliability estimate. As far its performance in the CFA was concerned, it was arguably the worst of the three single-factor models. The RMSEA value was slightly better than that for the PIL subscale, although well outside of Hu and Bentler's threshold of less than .06. Even by the earlier threshold of less than .08 Browne and Cudeck (1993) suggested for adequate fit, the value for the RMSEA was not acceptable. The values for the other fit indices were worse. (It should be noted however that these models are independent of each other and the values of the fit indices are not related to each other.)

The poor performance of these two subscales raises the question of the need for two preference scales for what, in fact, may be a single, but bipolar, construct. Put another way, a preference for individual learning and a preference for group learning may not be separate constructs, but rather two ends of a single continuum. In this case, there would only be a need to include one of the subscales in any proposed instrument, and there is the additional possibility that items from both could be included in the single scale, provided that items from

one of them were reverse coded. However, currently, this is an arguable case only, and there may be individuals who express a preference for both learning contexts, and in such a case a single bi-polar construct would not be plausible. Even if a single bi-polar construct is not plausible, it would still be possible to choose only one of the two subscales for inclusion in a new instrument, and this would have to be the PGL construct, given that the primary area of interest for this instrument is students' dispositions towards group work. This proposed elimination of the PIL subscale from any future adaptations of the instrument is also supported by the findings of Cantwell and Andrews' (2002) original study. In their study, a small but significant degree of correlation was found between the PIL and DGL subscales, whereas no significant correlation was found between the PIL and PGL subscales. For this reason, the elimination of the PIL subscale might not lead to a significant loss in the ability of the instrument to identify individuals who could find group work challenging. An additional advantage of the elimination of one the subscales would be the creation of a more abbreviated instrument. A more abbreviated instrument would help to limit the degree of fatigue experienced by respondents and make it more likely to provide better quality data.

The third single-factor model examined was the DGL subscale. As noted above, this scale had the lowest alpha value of the three subscales, as well as having some issues with the degree of skew in the scores for two of its items (Items 2 and 4). That being said however, the results of the CFA conducted on this subscale were exemplary, with the values of all of the indices above the cut-offs. Brown (2015) notes that for the CFI index, the closer the value is to 1 the better the degree of fit. Brown also comments that a value of 0 for the RMSEA index implies "perfect fit" (p. 72) between the model and the data. In addition, the chi-square value was small and, more importantly, non-significant, implying that that the dimensionality of the data does not differ significantly from that of the model. This finding reinforces the information from the other indices that the model is a good-fit. Finally, Mardia's coefficient was under 5, which indicates multivariate normality in the dataset.

The exemplary performance of the DGL subscale can be seen as a positive outcome of this study. As Cantwell and Andrews (2002) found, the largest contributing factor by far in learners' feelings of discomfort with group learning was their feelings of social anxiety. Anxiety, then, seems to be a construct that is quite salient to students and therefore more easily measured, and one profitable line of research which may follow from this study would be the further development of this subscale as a means to help identify students' who could be considered as "at risk" in terms of the increased prevalence of group-based approaches.

One limitation of this subscale is its short, four-item length. The results of the present study could be seen as a positive starting point for this scale to be expanded. The present version of the scale might have a rather narrow operational bandwidth regarding the range of possible feelings of discomfort that arise in group work because of its small number of items. The addition of further items representing other possible indicators of anxiety in groups could

help to expand the operational expression of the construct.

The first of such items might come from within the scale itself. Item 2, "*I sometimes feel nervous when I have to give my ideas or communicate within a group,*" seems to conflate two possibly discrete indicators, giving one's opinion and talking to others within the group. Fashioning two distinct items might provide clearer insights into the sources of learners' anxiety—simply interacting with others, or expressing an opinion in front of others. (This modification might also serve to reduce the degree of non-normality in the scale as Item 2 exhibited the third greatest skewness after Items 4 and 15.) Cantwell and Andrews (2002) found strong correlation between the DGL subscale, the Social Avoidance and Distress Scale and the Fear of Negative Evaluation Scale (Watson & Friend, 1969). The content of these two well-known scales could provide sources for the expansion of the scale. Items such as "*I try not to talk to people I don't know,*" or "*I am more likely to miss class when we are doing group work*" could be incorporated to operationalize aspects of avoidance behaviors, while items such as "*I worry about what other group members think of me,*" or "*I am afraid of making mistakes when working in a group*" might be included to reflect fears concerning possible negative evaluations by peers.

Once such items have been selected and added to the scale, this new version of the DGL scale could be submitted to exploratory analysis, and the results used in future development of the scale. The ultimate aim of this course of research would be the development of an instrument which was able to generate structurally valid and reliable scores in the Japanese EFL context. Moreover, the product of this development would have been informed by both theory and empirically derived evidence, allowing practitioners and researchers to utilize the instrument with a degree of confidence.

Conclusion

As detailed in the Introduction and Discussion above, cognitive and psychological factors appear to play a role in students' dispositions towards group work. Anxiety has been shown to be an influential factor in students' attitudes towards and performance in group learning situations (e.g., Cantwell & Andrews, 2002; Fantuzzo, 1989; Forrester & Taschian, 2010; Zhou, 2015), and therefore may be one of the most important issues in regards to group work. Placing students who have feelings of anxiety directly into group learning contexts without regard for these feelings could be detrimental not only to the individual students themselves, but also to the overall efficacy of the groups they are placed in. For educators to deal with this situation effectively, there needs to be a means of identifying such students before they are exposed to group-based learning approaches.

Up to now, one of the primary concerns with regards to anxiety in the EFL context has been that of the anxiety felt in communicating in a foreign language and this has emerged as an important research trajectory within the field of applied linguistics (e.g., Horwitz, Horwitz

& Cope, 1986; Matsuda & Gobel, 2004; Oxford, 1999; Woodrow, 2006). As group learning comes to be emphasized to a greater extent, and becomes more widely implemented in classrooms in general—and not just in the EFL context—attention will need to be given to the role that anxiety plays in group work.

With greater research on group-based learning approaches, it is possible that the anxiety involved in working in groups will come to be the predominant negative issue associated with the approach. In Japan in particular, with the phenomena of hikikomori, and other forms of social withdrawal, there may be a need to take special note of the anxiety of students whose dispositions are not geared towards group work. While the poor performance of the present adaptation of the FTGW instrument as a whole is disappointing, the very strong performance of the DGL subscale can be seen as a positive outcome of this study, and one which provides a possible avenue of exploration in any further work in this line of research. An interesting associated line of future research could be the examination of any possible correlations between this subscale and other instruments, such as Leary's (1983) Interaction Anxiousness Scale, that are more established in the Japanese context (Okabayashi & Seiwa, 1991).

Finally, in regards to the limitations of this study as a whole, it must be noted that this study deals with just one sample. Moreover, the sample is not a truly representative sample of the target population—Japanese EFL students—but a sample of convenience, which limits the generalizability of the results. Further studies sampling from the same population could help to ameliorate this limitation by giving a broader representation of the population of interest.

References

Angeli, C., & Tsaggari, A. (2016). Examining the effects of learning in dyads with computer-based multimedia on third-grade students' performance in history. *Computers & Education, 92*, 171-180.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, NY.: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Cantwell, R. H., & Andrews, B. (2002). Cognitive and psychological factors underlying secondary school students' feelings towards group work. *Educational Psychology, 22*(1), 75-91.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applica¬tions. *Journal of Applied Psychology, 78*, 98-104.

Costa, Paul T. and Robert R. McCrae (1992), *Neo PI-R and Neo-FFI Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Cowden, P. (2010). Communication and conflict: Anxiety and learning. *Research in Higher Education Journal, 9*, 1-9

Dörnyei, Z., & Malderez, A. (1997). Group dynamics and foreign language teaching. *System, 25*(1), 65-81.

Fan, X., & Thompson, B. (2001). Confidence intervals for effect sizes confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*(4), 517-531

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology, 81*(2), 173-177.

Forrester, W. R., & Tashchian, A. (2010). Effects of personality on attitudes toward academic group work. *American Journal of Business Education, 3*(3), 39.

Fushino, K. (2010). Causal relationships between communication confidence, beliefs about group work, and willingness to communicate in foreign language group work. *TESOL Quarterly, 44*(4), 700-724.

Gasaymeh, A. M. M., Kreishan, L. J., & Al-Dhaimat, Y. A. (2014). The Effects of Age, Gender, and Major on Measures of Jordanian Students' Attitudes towards Academic Group Learning in Computer Training Courses. *Journal of Education and Practice, 5*(15), 173-179.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale develop¬ment incorporating unidimensionality and its assessment. *Journal of Marketing Research, 15*, 186-192.

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. International *Journal of Educational Research, 39*(1), 35-49.

Gorvine, B. J., & Smith, H. D. (2015). Predicting student success in a psychological statistics course emphasizing collaborative learning. *Teaching of Psychology, 42*(1), 56-59.

Goudas, M., & Magotsiou, E. (2009). The effects of a cooperative physical education program on students' social skills. *Journal of Applied Sport Psychology, 21*(3), 356-364

Goudas, M., Magotsiou, E., & Hatzigeorgiadis, A. (2009). Psychometric properties of the Greek version of the Feelings toward Group Work questionnaire. *European Journal of Psychological Assessment, 25*(3), 204-210.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37*(4), 827-838.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125-132.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works?. *Change: The Magazine of Higher Learning, 30*(4), 26-35.

Kondo, D. S., & Yang, Y .L. (1994). Strategies for reducing social anxiety. *Communication Research Reports, 11*(2), 153-159.

Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of Personality Assessment, 47*(1), 66-75.

Lemonia, D., & Dimitris, G. (2017). Cooperative Learning as a Preference of Adolescent Students: The Case of Dance Activities. *Academic Journal of Interdisciplinary Studies, 5*(3 S1), 478.

Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41*(1), 585-634.

Maesin, A., Mansor, M., Shafie, L. A., & Nayan, S. (2009). A study of collaborative learning among Malaysian undergraduates. *Asian Social Science, 5*(7), 70.

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*(1), 21-36.

Ministry of Education, Culture, Sports, Science and Technology. (2003). *Regarding the Establishment of an Action Plan to Cultivate "Japanese with English Abilities."* Retrieved September 19th, 2017 from http://warp.ndl.go.jp/info:ndljp/pid/286794/www.mext.go.jp/b_menu/houdou/15/03/03033101/001.pdf

Ministry of Education, Culture, Sports, Science and Technology. (2008). Section 9 Foreign Languages. In *Improvement of Academic Abilities (Courses of Study).* Retrieved September, 19th, 2017 from http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2011/03/17/1303755_013.pdf

野崎秀正 [Nosaki Hidemasa]. (2012). 協同学習が仲間への質問行動に対する態度に及ぼす影響 [The Effects of Cooperative Learning on Students' Attitudes toward Asking Questions to Peers]. 宮崎学園短期大学紀要 [*Bulletin of Miyazaki Gakuen Junior College*], (5), 121-130.

野崎秀正 [Nosaki Hidemasa]. (2016). 授業中の協同活動に対する認識の学年差と達成目標の関連 [The Effects of Achievement Goals on Perceived Cooperation in Elementary and Junior High School Students]. 宮崎公立大学人文学部紀要 [*Bulletin of Miyazaki Municipal University Faculty of Humanities*], *23*(1), 53-60.

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory, 3*, 248-292.

岡林尚子 [Okabayashi Naoko], & 生和秀敏 [Seiwa Hidetoshi]. (1992). 対人不安感尺度の信

頼性と妥当性に関する一研究 [A study on the reliability and validity of Leary's social anxiety scale]. 広島大学総合科学部紀要. III, 情報行動科学研究 [Memoirs of the faculty of integrated arts and sciences, Hiroshima university. III. Studies in information and behavior science], 15, 1-9.

Oxford, R. (1999). Second language learning: Individual differences. In B. Spolsky (Ed.), *Concise Encyclopedia of Educational Linguistics* (pp. 552-560). Oxford, UK: Elsevier.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43-69.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33*(4), 448-457

White, F., Lloyd, H., Kennedy, G., & Stewart, C. (2005). An investigation of undergraduate students' feelings and attitudes towards group work and group assessment. *Research and Development in Higher Education, 28*, 618-623.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal, 37*(3), 308e328.

Zhou, M. (2016). The roles of social anxiety, autonomy, and learning orientation in second language learning: A structural equation modeling analysis. *System, 63*, 89-100.

Appendix - Feelings Towards Group Work Instrument (Cantwell & Andrews, 2002)

This questionnaire examines your preferences in relation to working and studying in groups.

Please read each statement and then indicate by circling the appropriate number, the degree to which you think the statement is true of you. If you believe the statement to be very true of you, circle the "5". If you believe the statement to be not true of you at all, then circle the "1". If you believe you are somewhere between these extremes, circle the "2", "3" or "4".

(1) I enjoy working within a group.
(2) I sometimes feel nervous when I have to give my ideas or communicate within a group.
(3) I understand information better after explaining it to others in a group.
(4) I often find it difficult to understand what the group task is.
(5) I like to work alone even when placed in a group.
(6) I prefer to work within a group rather than work alone
(7) I often have a strong feeling satisfaction when I become totally involved in a group achievement.
(8) It is important that other group members take responsibility for my learning as well.
(9) I usually make a strong personal contribution to group work..
(10) I am often afraid to ask for help within my group.

(11) I like group work more when we can make up our own groups.
(12) I do not like to study within a group.
(13) I can usually understand other group members' ideas.
(14) Even when groups are well organised, I don't believe they are a more effective way of using class time.
(15) It is best when each person helps each other within a group.
(16) I often think the work becomes too confusing when done in a group rather than individually.
(17) I rarely feel relaxed within a group.
(18) I sometimes feel let down by other group members.

Preference for Individual Learning subscale (PIL): Items 1, 5, 6, 12, 14, 16 and 18
Preference for Group Learning subscale (PGL): Items 3, 7, 8, 9, 11, 13 and 15
Discomfort in Group Learning subscale (DGL): Items 2, 4, 10 and 17

Psychometric Adaptation of a Japanese Version of the Feelings Towards Group Work Questionnaire for Use in the Japanese SLA Context

日本での第二言語習得における日本語版対グループワーク心理尺度の適用

セタキス　ラリー　ジョン

Abstract

This measurement study reports on the adaptation of the Feelings Towards Group Work questionnaire (FTGW; Cantwell & Andrews, 2002) into the Japanese population and second language acquisition domain. The original English version of the instrument was translated into Japanese, and this was then back-translated into English and compared to the original to ensure that the language used was equivalent. A data set (N = 307) was collected from university students at two universities in western Japan. Normality of test items was examined, and reliability estimates (Cronbach's alpha) of the three subscales that make up the instrument were calculated. Confirmatory factor analysis (CFA) was conducted to test the fit of the three-factor model hypothesized by the authors of the instrument, and the evidence was negative. Subsequent, and diagnostic, CFAs of the separate subscales indicated that the model fit for the scores on two of the subscales (Preference for Individual Learning and Preference for Group Learning) was unsatisfactory while the fit on the remaining subscale (Discomfort in Group Learning) was exemplary. The implications of these results for this adapted version of the FTGW questionnaire are discussed.

要約

本研究では、対グループワーク心理質問紙（FTGW; Cantwell & Andrews, 2002）の日本における第二言語習得分野での適用について報告する。原版の英語版尺度を邦訳し、これを再度英訳後に、原版と比較し言語の使用に相違が無いことを確認した。データ（N＝307）は西日本に所在する２大学の大学生から収集した。調査項目の正規性が確認され、尺度を構成する３つの下位尺度の信頼性（クロンバックα係数）を算出した。収集したデータが尺度の開発者の仮説である３因子モデルへ適合するかを確認するために確証的因子分析（CFA）を行ったが、適合は認められなかった。続いて検証のため行った下位尺度ごとの確証的因子分析では、２つの下位尺度（個人学習嗜好、グループ学習嗜好）のモデルへの適合は不十分であったが、１つの下位尺度（グループ学習での居心地の悪さ）においてはモデルへの適合は良好であった。これらの結果から、今回適用した FTGW 質問紙への示唆について考察した。