

熊本大学学術リポジトリ

Kumamoto University Repository System

Title	Examining the Psychometric Properties of the Groupwork Skills Questionnaire for use in the Japanese ...
Author(s)	Xethakis, Larry J
Citation	熊本大学社会文化研究, 17: 255-276
Issue date	2019-03-25
Type	Departmental Bulletin Paper
URL	http://hdl.handle.net/2298/42164
Right	

Examining the Psychometric Properties of the Groupwork Skills Questionnaire for use in the Japanese SLA Context

Larry J. Xethakis

Abstract

This measurement study reports on the adaptation of the Groupwork Skills Questionnaire (GSQ; Cumming, Woodcock, Cooley, Holland, & Burns, 2015) into the second language acquisition context. The target population of this study was Japanese university students. The first step in this adaptation process was the translation of the original English version of the instrument into Japanese. This was then back-translated and compared to the original to ensure that the language used was equivalent. Responses from university students at two universities in Western Japan ($N = 307$) comprised the data-set for this study. Normality of test items was examined, and reliability estimates (Cronbach's alpha) for the two subscales that make up the instrument were calculated. The fit of the two-factor model hypothesized by the authors of the instrument was tested using confirmatory factor analysis (CFA), and the evidence was negative. Subsequent, and diagnostic, CFAs of the two subscales comprising the GSQ indicated that the model fit for the scores on these subscales (Interpersonal Skills and Team Groupwork Skills) was less than satisfactory, though better than for the instrument as a whole. The implications of these results and possible suggested changes to this adapted version of the GSQ are discussed.

Introduction

Group-based learning approaches have been gaining in popularity in language learning classrooms in Japan (e.g., Fushino, 2010a, 2010b; Nosaki, 2012; Sakurai, 2015; Xethakis, 2016; etc.). One reason for this may be the shift in language teaching pedagogy taking place in Japanese schools. Increasingly, approaches to language teaching emphasize communication over more traditional grammar translation methods (Mitchell, 2017). Pair- and group work form an essential part of both Communicative Language Teaching and Task-based Language Teaching (Leeming, 2011), two of the more widely-used communicative methodologies. As these methodologies have come to figure more prominently in the language classroom, group-based learning approaches have also come to play greater roles.

Another reason for the increasing use of group-based learning approaches may be the priority given to them by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The use of group- and pair work are among the, so called, innovative approaches to learning that MEXT is emphasizing in secondary English language classrooms to promote students' abilities to communicate more fluently in English (MEXT, 2009; 2011). More generally, group-based learning activities, such as group discussions, debates and cooperative learning techniques, are being emphasized in all areas of the curriculum. In both cases, these approaches are seen to encourage active learning, which MEXT is championing as one means to help Japan react to a globalizing

society (MEXT, 2013; 2014).

Group-based learning has been shown to provide a wide range of academic, cognitive and social benefits. Johnson, Johnson and Smith (1998) note that approaches which emphasize cooperative over competitive or individual learning have been shown to enhance academic achievement, knowledge acquisition and higher-level reasoning in verbal, mathematical and even procedural tasks. Learners exposed to cooperative learning approaches also displayed higher motivation, greater persistence and more time on task. Moreover, such learners reported an increased sense of self-esteem, improved interactions with other learners, and a higher degree of social competence (Johnson, Johnson & Smith, 1998). In addition to the broad advantages conveyed by collaborative forms of learning, they can also provide benefits for language learning. Group members can form a pool of language resources, act as sources of motivation and bases of support for other learners (Dornyei & Malderez, 1997).

However, as Chapman and Van Auken (2001) state, “simply placing students into groups does not mean they will magically learn how to effectively work together,” (p. 118). The effectiveness of group-based learning approaches is influenced by a range of factors, including—personality traits (Rhee, Parent & Basu, 2013); anxiety triggered by social situations or a fear of negative evaluation (Cantwell & Andrews, 2002); personality conflicts, and the characteristics of the group task (McCorkle, Reardon, Alexander, Kling, Harris & Iyer, 1999); the existence of free-riders (Maiden & Perry, 2011); and previous negative experiences of group work (Hillyard, Gilleppe & Littig, 2010). All of these factors affect learners’ dispositions towards group work and their willingness to partake in it.

In addition to these factors, learners’ social skills also have a significant impact on the effectiveness of group work (Chapman & Van Auken, 2001; Johnson et al., 1998). Social and organizational skills are so important for successful groups that Johnson et al., (1998) include them as one of their five key elements essential for cooperation in groups (p. 31). Working in a group and dealing with the individuals involved can be a challenging situation. Contending with differing personalities and abilities among team members, managing conflict in a positive manner and supporting other group members, setting goals, defining roles and dividing up workload, these are among the skills that learners require to work successfully in groups. Placed in such a situation, without prior instruction or support, learners can easily become frustrated by their own “trial-and-error attempts to cope” (Schullery & Gibson, 2001, p. 12), with the result that the group process breaks down.

Teachers who employ group-based learning approaches, however, often neglect to instruct learners in these skills. One reason for this may be that they may not understand the importance of these skills, or they may be forced by time constraints to forego direct instruction in group-work skills. In addition, it is often assumed that learners already possess the skills needed to work cooperatively with others, or will learn them as an inherent part of the group-work process (Prichard, Stratford & Bizo, 2006).

However, this is not necessarily the case. Such skills do not often develop organically (Michaelson & Black, 1994; as cited in Prichard, Bizo & Stratford, 2006, p. 120) nor do students naturally understand how to work with others in a collaborative way (Johnson et al., 1998). Rather, placing learners in a group without ensuring that they possess such skills increases the chances that the

group will not function effectively (Chapman & Van Auken, 2001), thus depriving them of the benefits these approaches are able to convey.

For the advantages of group-based learning approaches to be realized, students need to be instructed in some form of basic group-work skills before undertaking group work (Johnson et al., 1998; Prichard, Bizo & Stratford, 2006). Training participants in group skills can lead to better planning, problem solving and communication in groups (Ellis, Bell, Ployhart, Hollenbeck & Ilgen, 2005). Moreover, skills training has been shown improve learners' attitudes towards group work, their degree of satisfaction with their group, and the group's cohesion (Prichard, Bizo & Stratford, 2006). Training in group-work skills not only improves the functioning of the group, but it also has a positive effect on individual learning as well (Prichard, Stratford & Bizo, 2006).

Group-work skills and group-work training have received a great deal of interest in recent years across all levels of education in Japan. (e.g., Inukai, 2005; Nakamura, 2013; Shobo, 2012). However, the two terms in Japanese, 'group-work skills' and 'group-work training', seem to embody a much broader conception of group-work skills and training than that described above. To the best of this author's knowledge, with respect to the Japanese literature, group-work skills tend to be seen as elements of more generic life or social skills (e.g., Inukai, 2005; Mizuno, Tazumi, Sumiya & Tago, 2008; Nakamura, 2013). The result of this is that training in these skills is seen as means to achieve a wide range of purposes including: addressing problematic behavior in the classroom (e.g., Tsumura, 2010); improving communication abilities (e.g., Inukai, 2003; Mori, 2016); and helping new students adjust to university life (e.g., Mizuno et al., 2008). The aims of research in group-work skills, as it is presently conceived in the Japanese context, are laudable, and without a doubt, important in helping achieve broader societal goals. These aims do not, however, appear to focus on the specific skills that learners require to make group-based learning approaches more effective.

Given the significant impact of group-work skills (as defined in this paper) on the success or failure of group-based learning approaches and their strong influence on learners' willingness to undertake group work, attention should be given to examining the role that such skills play in the effectiveness of group work in Japanese classrooms, as well as the efficacy of training learners in such skills. For this area of group-work research to be more fully realized, there is a need for an evidence-based measure which examines learners' group-work skills in the Japanese context. Such an instrument would not only allow researchers to explore this area with greater confidence in their findings, it would, at the same time, provide educators with a means to examine the effectiveness of explicit instructional interventions. Such a tool could also be employed to uncover specific student needs before such an intervention in order to provide more focused instruction. In addition, it could be used as a self-assessment tool for students involved in group work to encourage reflection on their own, or other group members', use of group-work skills in order to improve the effectiveness of the group. Such reflection on group processes, or group processing as this is often termed, comprises an important aspect of effectively functioning groups (Johnson et al., 1998).

One of the most extensively used instruments for measuring group-oriented social skills found in the English literature is the Teamwork Knowledge Skills and Abilities Test (Teamwork-KSA Test; Stevens & Campion, 1994; 1999). This instrument comprises 35 multiple-choice items. Each item describes a hypothetical team work scenario with four possible answers, only one of which is

considered the best course of action, and with the remaining three responses considered incorrect.

The Teamwork-KSA Test is based on a widely prevalent conceptual model of group-work skills which classifies the different actions that enhance team function into two broad categories: interpersonal skills and task management skills (Hobson, Stupeck, Griffin, Szostek & Rominger, 2014; Morgan, Glickman, Woodard, Blaiwes & Salas, 1986; Stevens & Campion, 1994). In this model, task management skills, such as setting goals, planning steps towards goals, monitoring and evaluating progress and providing feedback, are employed to help complete the group task, while interpersonal skills are focused on managing relationships between group members, including conflict resolution, exchanging information and providing social support.

While the Teamwork-KSA instrument is widely used for measuring group-work skills in industry and other areas, it does present a number of issues for use in the classroom. As a test of knowledge concerning group skills, it was not designed to be administered repeatedly. For this reason, it is of limited value in measuring changes in learners' skills due to interventions or providing feedback on these changes (Hughes & Jones, 2011). It also cannot be used as a self-reflective measure in student groups. More importantly, the validity of the instrument itself has been brought into question. O'Neill, Goffin, and Gellatly (2012) examined the factor structure of scores produced by the Teamwork-KSA Test using confirmatory factor analysis (CFA) and found that the proposed model did not fit the data sufficiently. Because of these deficiencies, the Teamwork-KSA test was rejected by this author as a possible candidate for adaptation and further study in the Japanese context.

Van Duzer and MacMartin (2000) developed an instrument for peer- or self-assessment of group-work skills. As this instrument is designed specifically to be used in assessing peers or oneself, it could also be used for assessing changes in learners' skills or for the purpose of reflection. In addition, this instrument is rather short, comprising eighteen items, including three multiple-choice items, four short-answer items and eleven Likert-scale items with four anchor points. The brevity of an instrument is a rather important attribute, as research has shown that longer instruments tend to induce fatigue in respondents (Rupp 2017; Fushino, 2010b, p.718). Notwithstanding these positive aspects, the instrument has a significant shortcoming. It is focused primarily on skills utilized in completing the group task and disregards the more social aspects of group work and interpersonal relationships. This focus would seem to make it a weak candidate for adaptation into the Japanese context under the requirements covered above.

The Comprehensive Assessment of Team Member Effectiveness (CATME; Loughry, Ohland & Moore, 2007) is another widely used instrument. The CATME comprises items assessing both interpersonal and task management skills, among a number of other constructs which include expectations, knowledge, skills and abilities. This comprehensive coverage comes with a significant drawback, however. The instrument is rather long, with the original version (Loughry et al., 2007) comprising 88 items, measured on a seven-point Likert-scale. As mentioned above, the fatigue induced by such lengthy instruments may interfere with the gathering of accurate data from respondents. Moreover, the CATME is designed to be backward looking; that is, designed to evaluate peers' or an individual's own contributions to a finished group project. Thus the instrument would be problematic for use in measuring the effects of instructional interventions.

A more recently developed instrument is the Groupwork Skills Questionnaire (GSQ) (Cumming,

Woodcock, Cooley, Holland, & Burns, 2015). This instrument is quite brief, comprising only ten items, and yet examines both interpersonal and task management skills. Each of the items covers the frequency with which respondents undertake specified actions when working in groups and are responded to on five-point Likert-scale. Because of its brevity and its focus on the frequency of actions performed in groups, it presents as quite suitable for use in measuring the levels of learners' group-work skills and in assessing the effectiveness of interventions, as well as student reflections on their own, or other group members' performance. Finally, and most importantly, the instrument has been shown to be a valid and reliable measure of both interpersonal and task management skills (Cumming et al., 2015).

The GSQ was developed and the psychometric properties of scores it produces were examined in four steps using responses from a total of 1653 university students with a wide range of majors. Scores from an initial sample of 672 respondents on a 46-item pilot version of the questionnaire were subjected to exploratory factor analysis (EFA). The results of this analysis suggested a two-factor solution, with the first factor comprising eight items relating to task management skills and the second factor comprising seven items related to interpersonal skills. These two factors were termed Team Groupwork Skills (TGS) and Interpersonal Skills (IS), respectively, by the authors. The two factors were found to be moderately correlated ($r = 0.54$, $p < 0.001$), which the authors took to mean that the two scales were measuring related but separate dimensions of group-work skills (Cumming et al., 2015).

Subsequent to the EFA, the authors carried out a CFA using a separate sample (725 respondents). A sub-sample of scores from these respondents ($N = 384$) was used to examine the fit of the factor structure suggested by the EFA. As a result of this process five items were removed from the questionnaire—three from the TGS subscale and two from the IS subscale. The resulting fit indices showed that the two-factor model had, according to the authors' interpretation, a decent fit (CFI = 0.92; TLI = .89; SMSR = 0.06; RMSEA = 0.08); although one would have to question the result for the TLI here, and the RMSEA is satisfactory by some standards (Byrne, 2001), but definitely not meritorious. These results were then confirmed using the scores from the remaining respondents ($N = 341$), with fit indices better than the above and indicating reasonable overall fit (CFI = 0.94; TLI = 0.93; SMSR = 0.06; RMSEA = 0.07).

Concurrent validity of the GSQ with two other instruments was established using scores from a third sample of 148. The first of these was a scale employed by Chapman and Van Auken (2001) to measure student's attitudes towards working in a group. The second was a scale to measure respondents' self-efficacy for group work, and was adapted from the Personal Efficacy Beliefs Scale (Riggs, Warka, Babasa, Betancourt & Hooker, 1994, as cited in Cumming et al., 2015). Positive correlations were found between scores on the self-efficacy for group-work scale and scores on both the TGS and IS subscales. However, a significant correlation (positive) was found only between scores on the attitudes towards group-work scale and the IS scale of the GSQ.

Test-retest reliability of the GSQ was established using a fourth sample of 108 respondents and was found to be acceptable. Finally, Cumming et al., (2015) report values for Cronbach's alpha as being over .7 for both subscales, with the exception of the first administration of the GSQ on the fourth sample, where the internal reliability of the TGS subscale was measured at .59.

The increased use of pedagogical approaches centered on group work, such as CLT and TBLT,

in English classrooms in Japan, and MEXT's push for teachers to employ group-based learning approaches to an even greater extent, together with the importance of group work skills for the effective implementation of group-based learning approaches, has created a need for evidence-based measurement of learners' group-work skills. Because of its brevity, its prospective utility for assessing the effects of instructional interventions and evidence for the reliability and validity of scores it produces, the GSQ (appended to the end of this paper) was chosen as the most suitable instrument for adaptation into the Japanese context. The present study, reporting on an examination of the psychometric properties of a Japanese version of the GSQ, represents an incremental step towards providing a secure empirical foundation for research in this area.

Methodology

Below, the methodology of the study is reported in three sections. First, the characteristics of the GSQ are reported, followed by a description of the participants and an explanation of the procedure followed in the collection of the data. Finally, the analytical processes employed—the statistical and psychometric techniques utilized in the process of examining the scores in the data set for this study—are outlined.

Instrument

The GSQ (appended to the end of this paper) has its ten items divided equally between two hypothesized subscales—Interpersonal Skills (IS, Items 1, 3, 5, 7 and 9), and Task Groupwork Skills (TGS, Items 2, 4, 6, 8 and 10). Each of the items asks the frequency with which respondents undertake specified actions when working in groups. Responses are recorded on a five-point Likert-scale, with 1 being semantically anchored to *never* and 5 to *always*. Scores for each of the subscales are computed separately, and there is no composite score for the entire instrument.

As mentioned in the Introduction above, the GSQ was originally intended for use in an L1 English speaking university population. Therefore, the first step in adapting the instrument for use in the Japanese EFL context was to translate the instrument from English into Japanese. This was done following the guidelines presented by the International Test Commission (Hambleton et al., 2005). An initial translation was undertaken by the author with the assistance of a native Japanese speaking professor. This translation was given to two native English-speaking university instructors, both of whom are fluent in Japanese, for the purpose of backtranslation. All of the translators involved had some experience in the field of testing.

A small number of discrepancies in the reverse translations led to a revised version of the translation and after consensus on the wording of the items was reached, the translation was deemed ready for use in the Japanese EFL context.

Participants and Data Collection Procedure

There were a total of 363 responses collected from students at two universities in Western Japan. 32 of these were removed from the data set for missing responses on one or more items. The removal of these responses was considered not to have a negative impact on the overall data set as there was, by inspection, no pattern to the missing responses. In addition to the missing

responses, a further 24 responses were removed for displaying obvious patterns (such as all 1s). This left 307 responses remaining in the data set, and the analysis in this paper is based on this data.

The age of the respondents varied from 17 to 25, with a median age of 20 years old. Among the respondents there were 236 males and 70 females (one respondent did not provide a gender). The largest number of respondents came from students in the field of computer science and electrical engineering ($n = 128$), followed by those in civil and environmental engineering ($n = 62$). The number of respondents from other academic departments is as follows: applied chemistry and biochemistry ($n = 21$), business administration ($n = 20$), medical care and welfare engineering ($n = 19$), law ($n = 18$), tourism management ($n = 13$), animal science ($n = 11$), plant science ($n = 8$), and bioscience ($n = 7$).

In addition to their age, gender and department, respondents were asked about their experience with group work in junior high school, high school and university English classes. Responses to these questions were given on a seven-point Likert-scale, with 1 semantically anchored to *never* and 7 to *almost always*.

Participation in the survey was completely voluntary, with students being informed by the administrator of the survey that they need not take part. In addition, a form asking for participants' informed consent, clearly stating in Japanese that those not wishing to participate could do so merely by leaving the form blank, was included at the beginning of the survey form. While no set time was given within which to complete the survey, most participants did so within 10 minutes.

Analytical Procedure

Initially, participants' data (including their scores on the GSQ as well their age, gender, department and group work experience) was entered into a Microsoft Access 2016 database. This data was then imported into IBM/Statistical Package for the Social Sciences (SPSS) software (Version 21) in order to calculate descriptive statistics and reliability estimates (Cronbach's alpha) for the dataset. AMOS (Version 21) was employed to carry out the confirmatory factor analysis (CFA) on the scores. The descriptive statistics generated by the scores in the dataset were examined first with an eye to their univariate normality (i.e. skew and kurtosis). Subsequent to the determination of the degree of skew and kurtosis found in the scores of each of the items, reliability estimates of the two subscales—IS and TGS—were calculated. Finally, CFAs were conducted on: (1) the GSQ taken as whole; and (2) both of the instrument's subscales considered independently.

For the purpose of determining the degree of skew and kurtosis found in the scores, the critical ratio was calculated by taking the skew and kurtosis values for each of the items and dividing these values by their respective standard errors (See Table 1 for these values, and Table 2 for the respective critical ratios for each item). A minimum acceptable value of 3.0, as well as a stricter criterion of 2.0, and a meritorious criterion of equal to or less than 1.0, were employed as interpretive criterion to determine the degree of skew and kurtosis. These values were stipulated by the author in advance.

Reliability estimates (Cronbach's alpha) were computed following the recommendations of Fan and Thompson (2001), who suggest that these computations include estimations of confidence

intervals (95%). For the interpretation of alpha, Nunnally and Bernstein's (1994) criterion—a value of .70 for the reliability of the scale—was adopted.

For the central analysis of this study, the CFA, four fit indices—the Tucker-Lewis index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean squared residual (SRMSR)—were utilized in conjunction with the chi-square. The chi-square as an indicator of model fit is sensitive to the size of the sample, and thus it tends to over-reject models when the sample size is smaller. Employing the four fit indices in combination with the chi-square is one means to overcome this tendency. Hu and Bentler (1999) have recommended cut-off values for each the four fit indices (TLI and CFI >.95; RMSEA <.06; SRMSR <.08), which are used in conjunction to evaluate the fit of the model, and these values have been adopted to evaluate the fit of the hypothesized model for the GSQ.

Results

The results section of this study consists of three sections. The properties of the sample and the distribution of the scores on each item are presented in a description of the descriptive statistics, which include the means, the standard deviation, and skew and kurtosis for each item. This represents the first of the three sections. The reliability estimates (Cronbach's alpha and the associated confidence levels for alpha) are presented in the section section. The results of the CFAs are presented in the final section. These include the results of the CFA on the instrument as a whole (correlated model), as well as on each of the two subscales (taken as independent uncorrelated models as there is only one factor in each case).

Descriptive Statistics

The descriptive statistics of the dataset, with respect to the distribution of the scores on the items of the GSQ, are shown below in Table 1. Item 3 had the highest mean among the items at 3.73, while Item 1 had the lowest, at 2.65. The range of the standard deviations for the items was between 0.971 (Item 2) and 0.817 (Item 3).

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

Item	Mean	SD	Skewness		Kurtosis	
			Statistic	Std. Error	Statistic	Std. Error
01	2.65	0.844	0.446	0.139	0.280	0.277
02	2.73	0.971	0.500	0.139	-0.297	0.277
03	3.73	0.817	-0.487	0.139	0.588	0.277
04	3.16	0.943	-0.118	0.139	-0.276	0.277
05	3.67	0.863	-0.386	0.139	0.137	0.277
06	2.91	0.954	0.359	0.139	-0.371	0.277
07	3.36	0.960	-0.139	0.139	-0.446	0.277
08	2.99	0.943	0.144	0.139	-0.407	0.277
09	3.24	0.833	0.207	0.139	0.132	0.277
10	2.92	0.911	0.234	0.139	-0.075	0.277

The results from the calculation of the critical ratio for each of the item's skew and kurtosis is presented in Table 2. As mentioned above in the Methodology section, this value was calculated by taking the respective values for skew and kurtosis for each item and dividing them by their respective standard errors. The resultant values were then compared to the aforementioned criteria (1.0, 2.0 and 3.0). Values less than 2.0 have been left unmarked, while those failing to meet the 2.0 standard are marked with one asterisk. Values which did not meet even the more relaxed criteria of 3.0 are marked with two asterisks.

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

Item	Skewness	Kurtosis
	Calculated Value	Calculated Value
01	**3.21	1.01
02	**3.60	1.07
03	**3.50	*2.12
04	0.85	1.00
05	*2.78	0.49
06	*2.58	1.34
07	1.00	1.61
08	1.04	1.47
09	1.49	0.48
10	1.68	0.27

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

Examining the values for the skew of the each of the items shows that fully half of the items fell below the stricter 2.0 threshold, with three of these items (4, 7 and 8) falling below or just slightly above the meritorious value of 1.0. Two of the remaining five items fell below the 3.0 threshold while three of the ten items on the instrument exceeded this threshold. The items performed better in terms of kurtosis. Nine of the ten items fell below the strict 2.0 threshold, with the remaining item (Item 3) falling just 0.12 above this threshold. The extent to which the degree of non-normality found in some of the items was either similar or dissimilar to that for items in the original version of the instrument cannot be determined, as Cumming et al. (2015) did not report such information. This, therefore, also means that it is not possible to determine whether the non-normality evidenced in the items reported above is specific to the Japanese population and this particular adaptation, or whether such problems might be observed across other populations; in other words, being a problem which is relatively population invariant.

Reliability Estimates

Table 3 shows the reliability estimates (Cronbach's alpha) and respective confidence intervals for the scores of both of the hypothesized subscales comprising the GSQ (Cumming et al, 2015). Scale means and standard deviations are also represented.

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

Subscale	Cronbach's alpha	95% Confidence Intervals for Cronbach's alpha		Scale Mean	SD for Scale
		Lower Bound	Upper Bound		
IS	.745	.697	.788	16.65	3.043
TGS	.686	.627	.738	14.71	3.145

The IS subscale value for reliability was reasonable, with the value for alpha exceeding the .70 threshold (discussed above in the Methodology section) and the lower bound of the 95% confidence interval only just missing this threshold. The TGS scale could be seen as slightly problematic. The value for alpha on this subscale fell marginally below the .70 threshold and moreover, the lower bound of the 95% confidence interval fell rather below the threshold of .70 cited above. Cumming et al. (2015) reported values of greater than .70 for alpha on both subscales in three of the four samples in their study. In the fourth sample, the reliability estimate for the IS subscale (.73) was above the threshold, however, that for the TGS (.59) fell rather far below the threshold value. Moreover, the reported values for the reliability estimate in sample 1 (which was used for the EFA) also showed this tendency. The values reported for that sample were: IS .81 and TGS .72. The reoccurrence of this trend in the present study—the reliability of the TGS subscale estimated as lower than that of the IS subscale—may be evidence of an inherent and population invariant weakness within the items of the TGS subscale. However, Cumming et al. (2015) note that the removal of items from the subscale did not alter the reliability estimate in the scores from Sample 4 and thus the subscale was left unaltered in the original study.

Confirmatory Factor Analysis

For the purpose of testing the two-factor structure of the GSQ as hypothesized by Cumming et al. (2015), a CFA was carried out on the scores from the data set described above, and with a model so specified; that is, a two-factor correlated model. In further analysis, both of the instruments' component subscales were treated as independent single-factor models, and these two models were directly tested using CFA as well. In order to evaluate the fit of the models to the scores in the data set, four fit indices were employed in conjunction with the chi-square. The chi-square is rather sensitive to sample size, and it is to help overcome this limitation that fit indices are employed, and these are typically used in combination with each other, in order to determine model fit. In this study, empirically derived cut-off values for the four indices employed have been adopted following the recommendations of Hu and Bentler (1999).

The four indices utilized in examining each of the models considered in this study were also recommended by Hu and Bentler (1999). They were specifically recommended to be used in triangulation, and include the RMSEA, the SRMSR, the TLI, and the CFI. Brown (2015) notes that these four indices are among the most commonly reported in applied research. The RMSEA is not sensitive to sample size, as is the chi-square, and moreover, it favors model parsimony, or in other words, it rewards models with a simpler structure, and thus more complex models will appear to have poorer fit (Harrington, 2009). The SRMSR is an absolute fit index and evaluates models based

on differences between the correlations observed in the data set and those predicted by the model (Brown, 2015). For both the RMSEA and the SRMSR, the smaller the value, the better the result is considered to be. The TLI and CFI assess the hypothesized model relative to the baseline model (termed the independence model). For both of these indexes, higher values suggest better fit. The CFI has a range of 0 to 1, however the TLI, having a value that is non-normed, can yield values outside the 0 to 1 range (Brown, 2015).

Table 4 presents the values that were calculated for the goodness-of-fit indices employed in analyzing each of the models considered in this study. Also shown are the cut-off values suggested for each of the indices by Hu and Bentler (1999).

Table 4. Comparison of Goodness-of-fit Indicators for GSQ Two-Factor Model and Subscale One-Factor Models

Index	Cut-off value	Scale / Subscale		
		GSQ	IS	TGS
TLI	(>.95)	.804	.842	.879
CFI	(>.95)	.852	.921	.940
RMSEA	(<.06)	.104	.131	.092
SRMSR	(<.08)	.0662	.0529	.0440
Mardia'scoefficient	(< 5)	25.333	12.822	8.249
Chi-square	(value)	146.279	31.420	18.077
	(probability level)	.000	.000	.003

Two-Factor Model

The first model considered is that hypothesized by Cumming et al. (2015) as a result of the process of EFA and CFA described above in the Introduction. This model comprises two factors, which the authors suggest measure respondents' proficiency in utilizing group-work skills. These factors are stipulated to align with the IS subscale (Items 1, 3, 5, 7, 9) and the TGS subscale (Items 2, 4, 6, 8, 10); and with these factors allowed to correlate. The model had 55 distinct sample moments, 21 distinct parameters to be estimated, and 34 degrees of freedom, meaning the model was overidentified.

Mardia's coefficient was used to assess the multivariate normality of the data. A value over five indicates multivariate non-normality for this measure. The value for this particular model was 25.333, indicating multivariate non-normality in the data. It is interesting to note that Cumming et al. (2015) also reported a degree of multivariate non-normality in the scores from the data sets used for the CFA in the original study (subsample 2.1 = 43.65; subsample 2.2 = 40.96).

The results from the calculations of the fit indices were as follows (Hu and Bentler's cut-off values given in parentheses): TLI .804 (>.95), CFI .852 (>.95); RMSEA .104 (<.06); SRMSR .0662 (<.08). The chi-square value was 146.279 with a probability level of .000. Taken together, these four indices and the chi-square indicate that the model hypothesized by the original authors is problematic in the present adaptation of the instrument and lacks sufficient fit.

Single-Factor IS Model

This model comprises only one of the two subscales, the IS subscale (Items 1, 3, 5, 7 and 9), in the model originally hypothesized by Cumming et al. (2015). It was specified as a single unidimensional factor. The model had 15 distinct sample moments, 10 distinct parameters to be estimated, and 5 degrees of freedom, and thus was overidentified.

Mardia's coefficient for this model was 12.822, indicating multivariate non-normality in the data. The results for this model were as follows (Hu and Bentler's cut-off values given in parentheses): TLI .842 (>.95), CFI .921 (>.95); RMSEA .131 (<.06); SRMSR .0529 (<.08). The chi-square value was 31.420 with a probability level of .000. These results indicate that the data set fits the model for this subscale insufficiently and, therefore, this single-factor model is problematic in the present adaptation of the instrument.

Single-Factor TGS Model

As with the results reported directly above, this model also comprised one of the subscales in Cumming et al.'s (2015) hypothesized model, the TGS subscale (Items 2, 4, 6, 8 and 10). The model had 15 distinct sample moments, 10 distinct parameters to be estimated, and 5 degrees of freedom, and thus this model was also overidentified.

Multivariate nonnormality in the data was indicated by the value for Mardia's coefficient for this model, 8.249, however, the value was just over the threshold and is less problematic than for the other two models, and especially when compared with the two-factor-model. For this model, the results of the fit indices were as follows (Hu and Bentler's cut-off values in parentheses): TLI .879 (>.95), CFI .940 (>.95); RMSEA .092 (<.06); SRMSR .0440 (<.08). The value of chi-square was 18.077 with a probability level of .003. These results indicate that there is an insufficient degree of fit between the model and the dimensionality of the scores in the data set. Therefore, in the present adaptation of the instrument, this single-factor model is also problematic, though better than for the other two models tested.

Discussion

This study reports an examination of the dimensionality of the scores produced by a version of the GSQ adapted for use in the Japanese EFL context. The GSQ has been shown to be a valid and reliable instrument for measuring group-work skills in an English-speaking context and within the population of university students (Cumming et al., 2015). The specific aim of this paper was to determine if the structure of the instrument proposed by Cumming et al. (2015) would be equally valid for scores generated in the Japanese university student population. Such a determination is important for an evidence-based judgement as to whether the instrument is useful to measure the group-work skills of Japanese university students. If the model were to be found plausible in this context and with this population, the GSQ could become a practical instrument for assessing the effectiveness of pedagogical interventions, as well as providing researchers with a tool for exploring the area of group-work skills with greater confidence.

However, the results from the central analysis of the paper, the CFA, yielded a less than

satisfactory result for the two-factor structure proposed by Cumming et al. (2015). Following these results, additional diagnostic CFAs were undertaken on both of the sub-scales that comprise the instrument. These analyses revealed that the hypothesized structure of both subscales, when taken independently as a single-factor unidimensional model, exhibited poor-fit with the scores in the data set as well.

While the uniformly negative results from the CFAs described in this study may be disappointing in terms of confirming the structure of the adapted instrument in relation to the dimensionality of the scores in the data set, the results of the analyses presented in this paper can serve a valuable function as an empirically grounded aid in further adaptation of the GSQ for the Japanese EFL context. In this respect, the fundamental issue concerns the changes that would be required in the items comprising the GSQ. Any necessary changes would lead to a new version of the instrument whose structure could then be tested against the dimensionality of the scores in a new data set.

One area of concern was the degree of skew found in some of the items. Five of the ten items in the instrument (Items 1, 2, 3, 5 and 6) exhibited a degree of skew that exceeded the 2.0 threshold stipulated beforehand, with three of these exceeding even the more relaxed 3.0 threshold (Items 1, 2, and 3; See Table 2). When the scores on an item are skewed to a significant degree, they are grouped primarily around one or the other end of the scale. This clustering of scores limits the degree of variability expressed in the responses. It is this variability that provides information concerning individual differences and their influence on behavior, and thus a strong positive or negative skew represents a reduction in the variability of the responses to an item and as a consequence a loss of such information.

One possible solution to this loss of information would be to expand the Likert scale used in the instrument from a five-point scale to a seven- or even ten-point scale. By expanding the scale to such degree, the scale would ideally become more sensitive to respondents' use of group-work skills. However, it is debatable whether respondents would be able to make such fine-grained distinctions, particularly in this case, where the scale is semantically anchored to language expressing the frequency with which respondents perceive themselves undertaking the action described in the item, rather than the extent to which respondents feel that the content of the item applies to them as is more commonly the case for these scales. If respondents have difficulty in making such fine-grained distinctions, and this would seem to be a possibility considering the idiosyncratic variations that can occur in the use of adverbs of frequency such as those that are utilized as semantic anchors in this instrument, then expanding the scale, rather than providing more information, would simply introduce a greater degree of 'noise' (i.e. error) into the responses.

Another possible change to the instrument could be the simple removal of those items that exhibited excessive skew and/or kurtosis. However, an objection to the removal of any particular item from the instrument would be that it is already quite brief, comprising only ten items, and doing so might merely serve to restrict the operational bandwidth of the subscale, and therefore the instrument as a whole. This would result in a reduction, rather than an increase, in the amount of information provided by the instrument.

A third option to be explored in the adaptation process would be an examination of the wording and content of the items that make up the instrument. The purpose of this examination would be

to help determine what aspects of those items might be limiting or influencing the range of responses on the particular item. Of the two subscales that comprise the GSQ, the IS subscale—concerned with actions aimed at maintaining or encouraging interpersonal relationships within a group—seemed to exhibit a number of issues in this regard.

The IS subscale included items with the greatest degree of non-normality, with three of the five items on the subscale (Items 1, 3, and 5) possessing a large degree of skew. Item 5 surpassed the stipulated 2.0 value by a considerable amount, while Items 1 and 3 exceeded the more relaxed value of 3.0 (See Table 2). Interestingly, these two items had opposite skews. Item 1 had a positive skew, and therefore a mean lower than the middle point of the scale. The skew of Item 3, on the other hand, was strongly negative, and thus this item had a mean above the midpoint. In fact, these two items had the lowest and highest means respectively (Item 1, $m = 2.65$; Item 3, $m = 3.73$) among the ten items on the GSQ.

In addition to the degree of skew, Item 3, “*When working in groups, I tend to be sensitive to the feelings of others*”, was the only item whose kurtosis exceeded the 2.0 value, at 2.12, and thus the responses to this item not only had a negative skew but also a relatively high peak. Out of the 307 respondents in this study, 149 of them, or 48.5%, responded with a 4, or *quite often*. Moreover, only 16 of them, or 5.2%, responded *not very often* or *never* in considering the frequency with which they perform the actions implied by the wording of this item. This means that almost 95% of respondents saw themselves acting in this way at least somewhat frequently when they are engaged in group work, and thus the item is picking up that the behavior is relatively invariant, and therefore, not a likely a point of significant individual difference.

One possible reason for the degree of clustering of responses with respect to this item might be found in the nature of most Japanese students. It is arguable that, as a culturally-driven behavioral proclivity, Japanese students tend to exhibit a higher degree of awareness of other individuals, particularly when working as a member of a group. For this reason, the actions implied by the wording of Item 3 might be seen as obvious behaviors when working with others. This tendency may not be expressed as strongly by Western students, with their inclination towards individuality, and thus students in an America, or other countries whose culture has a stronger emphasis on the individual, might tend to have a broader range of responses to the content of this item. Item 5, “*When working in groups, I tend to show that I care about my group members*”, which had the second highest mean of all the items on the GSQ ($m = 3.67$) and analogous response pattern—92.5% of respondents perceived themselves as acting in this manner sometimes or with greater frequency—might also be understood in terms of this hypothesis.

Another possible source of ‘noise’ or error in the data set gathered in this study may be the terminology of the items comprising the IS subscale. The terms used in this subscale, for example, “*provide emotional support to my group members*” (Item 1), “*be open and supportive when communicating with others*” (Item 7), or “*be there for other group members when they need me*” (Item 9), are most likely familiar to Western university students because of the common use of such terminology in popular psychology, and it can be presumed that they would also be familiar to educators in Japan with some understanding of psychology and interpersonal dynamics (such as those who worked in the translation process of this instrument). Because of this familiarity, there would be a number of attendant actions that may come to mind when these terms were considered by either of these

two groups, and thus it might be relatively straightforward for them to consider the frequency with which they performed such actions in a group context. On the other hand, Japanese university students, who most likely have not been exposed to the ideas and concepts of Western popular psychology to such an extent, might find these terms and phrases as unfamiliar and rather vague. This ambiguity, or the inability to consider what kinds of actions might comprise these concepts, could create uncertainty or even difficulty for respondents.

The pattern of responses to Item 9 might provide evidence to support this supposition. On this particular item, a majority of respondents (161 of 307, or 52.1%) chose 3, or *sometimes*. On a five-point scale, the midpoint, or 3, is often seen as the 'neutral' response, meaning that respondents are choosing not to make a definitive positive or negative choice. The phrase "*be there for someone*" has a positive implication of aiding or supporting other group members if they are facing a difficulty (and therefore 32.5% [100 out of 307] responded that they acted in this way *often or always*), however just what kind of actions, or what form of support would be provided may have remained unclear to many respondents. For this reason, they may have felt that they had in fact given some kind of support to other group members and yet were unable to definitively determine how often they performed such actions. Because of this a majority opted for the neutral response of *sometimes* on this item.

In a case such as that outlined above, this subscale might present an instance where expanding the Likert scale employed in the instrument from a five-point to a six-point scale might be warranted. In such circumstances, moving to a six-point scale would remove the neutral response and thereby force respondents to make a non-neutral response. This may assist with creating a greater distribution of responses to the items on the subscale.

Another means of improving the range of response for this subscale might be the use of focus groups of students to clarify those items whose meanings are particularly vague or difficult to understand and the reasons why focus group members felt this way. A similar undertaking was conducted by Rupp (2016) in examining the content of the Kambara Locus of Control instrument. An additional benefit of convening focus groups could be the development of a number of concrete examples of actions that would be included in the concepts employing terminology from popular psychology, such as '*being there for someone*' or '*being open and supportive when communicating*'. These focus group generated examples would presumably make the context of the items on the IS subscale less ambiguous and more easily understood by respondents who are less familiar with Western popular psychology. These examples could then be included in a revised version of the scale, whose properties could again be tested in the Japanese university student population.

The second subscale of the GSQ, the TGS subscale, which deals with the management of group tasks, performed slightly better than the IS subscale in terms of the normality displayed by its items. Of the five items on this subscale, only two items, Items 2 and 6, exhibited skew that was above the stipulated value of 2.0 (see Table 2). However, it should be noted that Item 2 had the largest skew of all the items of the GSQ, at 3.60. The skew of Item 2, "*When working in groups, I tend to remind the group how important it is to stick to schedules,*" was strongly positive, with a mean of 2.73, implying that respondents performed this action with relative infrequency. Thirty-nine percent of respondents, or 120 out of 307, chose *not very often* when considering the frequency with which they tended to perform this action during group work. A possible source for the pattern of

responses on this item, and the other items on this subscale as well, might be differences inherent in the organization and functioning of groups, such as the apportioning of group roles or taking initiative within the group. The phrasing of this item in particular is rather direct and suggests that the respondent is taking charge and directing the actions of other group members. While individual group members in a Western tertiary educational context might feel it appropriate for them to step forward and speak out or take charge of some aspect of the group's activities, such actions might not be so easily undertaken by Japanese university students.

Cutrone (2010) in discussing differences in communication styles between Japanese and English-speaking university students, cites Ellis (1991) who observes that Japanese students "may lack the politeness strategies needed to successfully perform facing-threatening speech acts such as invitations and requests" (p. 116). Reminding group members to stick to a schedule could be seen as a rather direct request of other members of the group and thus be possibly problematic for a large number of students. In addition to this proposed lack of communication strategies which hampers making requests of others, the high degree of communication apprehension found in Japanese students reported by McCroskey, Gudykunst and Nishida (1985) may also play a role. Individuals who are reticent to communicate in general, would most likely find it difficult to make requests or direct the actions of fellow group members. As Nozaki (1993) notes quietness and passivity are among the traits Japanese feel that good learners should possess. For this reason, Japanese learners might be hesitant to take initiative or take charge of a group, particularly in an English class. King and Smith (2017) observe that the classroom in general and the language classroom in particular, represent a "highly public social performance situation" (p. 99) for Japanese learners, which leads them to reduce their levels of verbal interaction. For these reasons, for one member to stand out and begin to organize aspects of the functioning of the group might occur less frequently in Japanese language classrooms than it might in classrooms in other cultures.

One piece of evidence which might lend credence to this hypothesis is the differences with which respondents perceive themselves undertaking the types of actions described by the two subscales that comprise the GSQ. Respondents perceived themselves using significantly less task management skills (Scale mean = 14.71) than interpersonal skills (Scale mean = 16.65). There is an almost two-point difference between the means of the two subscales. This implies that respondents undertook task management actions significantly less frequently than actions aimed at maintaining interpersonal relationships in the group; and this was confirmed statistically ($t(306) = 12.31, p < 0.000$). Cumming et al. (2015) found a similar trend in the data from the original study—the scale mean for the TGS subscale was 17.6, while that for the IS was 19.7. However, the difference between the means the subscales in the data set of Cumming et al. and the means of the subscales in the data set in this study is more than three points on each respective subscale. This large difference between the two populations would seem to reinforce the supposition that Japanese learners are more reticent to undertake task management actions, and even actions aimed at strengthening interpersonal relationships, in a group work setting.

If this finding holds true, it would underscore the need for Japanese learners to be provided with some form of group-work skills training before engaging in group work. Such interventions would allow for more effective group functioning, and thus a better realization of the benefits of group-based learning approaches.

Conclusion

This study has pointed out the need for learners to employ group-work skills and also the need for learners to be trained in the use of these skills in order for group-based learning approaches to become the effective learning tools they have been shown to be in the literature. For further research into the role that group-work skills play in group-based learning approaches, there is a need for an evidence-based instrument to measure learners' use of such skills. Furthermore, such an instrument would also be valuable to educators as a means of measuring the effectiveness of interventions aimed at encouraging learners' use of these skills when working in groups.

In this regard, the less than satisfactory performance of the GSQ examined in this study is disappointing. Nonetheless, the data gathered in the study can be of use in suggesting possible adjustments that could be made in further adaptation of the instrument for use in the Japanese EFL context. Once such changes have been made, the revised instrument could then be tested against a new dataset, and the fit of the model re-examined. Pursuing such an approach to further adapting the GSQ is arguably imperative, because there are no substitute instruments available in the Japanese context at present which have well-documented empirical support.

Finally, it must be noted that one limitation of this study is that it deals with only a single sample, and a sample that is one of convenience rather than a true representative sample of the target population. For this reason, the findings of this study may not be generalizable to the target population as whole. One means of overcoming this limitation would be further research examining the characteristics of a broader sample of Japanese university EFL students.

References

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY.: Guilford Press
- Byrne, B. M. (2001). *Structural equation modeling with AMOS*. London: Lawrence Erlbaum Associates.
- Cantwell, R. H., & Andrews, B. (2002). Cognitive and psychological factors underlying secondary school students' feelings towards group work. *Educational Psychology, 22*(1), 75-91.
- Chapman, K. J., & Van Auken, S. (2001). Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education, 23*(2), 117-127.
- Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment & Evaluation in Higher Education, 40*(7), 988-1001.
- Cutrone, P. (2010). Helping Japanese ESL/EFL learners overcome difficulties in intercultural communication. *Journal of the Faculty of Global Communication, University of Nagasaki, 11*, 55-63.
- Dörnyei, Z., & Malderez, A. (1997). Group dynamics and foreign language teaching. *System, 25*(1), 65-81.
- Ellis, A. P., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology, 58*(3), 641-672.
- Ellis, R. (1991) 'Communicative competence and the Japanese learner. *JALT Journal, 13*(2), 103-127.
- 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
- Fushino, K. (2010a). Let's try cooperative learning group projects! In A.M. Stoke (Ed.), *JALT2009 Conference Proceedings*. (pp. 359-369). Tokyo: JALT
- Fushino, K. (2010b). Causal relationships between communication confidence, beliefs about group work, and willingness to communicate in foreign language group work. *TESOL Quarterly*, 44(4), 700-724.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Harrington, D. (2008). *Confirmatory factor analysis*. New York, NY: Oxford University Press.
- Hillyard, C., Gillespie, D., & Littig, P. (2010). University students' attitudes about learning in small groups after frequent participation. *Active Learning in Higher Education*, 11(1), 9-20.
- Hobson, C., D. Strupeck, A. Griffin, J. Szostek, and A. S. Rominger. 2014. Teaching MBA students teamwork and team leadership skills: An empirical evaluation of a classroom educational program. *American Journal of Business Education*, 7, 191-212.
- 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.
- Hughes, R. L., and S. K. Jones. (2011). Developing and assessing college student teamwork skills. *New Directions for Institutional Research*, 2011(149), 53-64.
- 犬飼己紀子 [Inukai Mikiko]. (2003). コミュニケーションのスキルアップをねらったグループワーク・トレーニングの実践 [Using group work training to improve learners' communication skills]. *上田女子短期大学紀要 [Bulletin of Ueda Women's Junior College]*, 26, 11-21.
- 犬飼己紀子 [Inukai Mikiko]. (2005). 人間関係トレーニングとその効果：中学校1学年生徒へのレクレーション・GWT 授業実践から [Effect of human relationships training: Group learning with junior high school freshman]. *上田女子短期大学紀要 [Bulletin of Ueda Women's Junior College]*, 28, 19-30.
- 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.
- King, J., & Smith, L. (2017). Social anxiety and silence in Japan's tertiary foreign language classrooms. In C. Gkonou, M. Daubney & J-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications*, (pp. 91-109). Bristol: Multilingual Matters
- Leeming, P. (2011). Japanese high school students' use of L1 during pair-work. *International Journal of Applied Linguistics*, 21(3), 360-382
- Loughry, M. L., Ohland, M. W. & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, 67(3), 505-524.
- Maiden, B., & Perry, B. (2011). Dealing with free-riders in assessed group work: Results from a study at a UK university. *Assessment & Evaluation in Higher Education*, 36(4), 451-464.
- McCorkle, D. E., Reardon, J., Alexander, J. F., Kling, N. D., Harris, R. C., & Iyer, R. V. (1999). Undergraduate marketing students, group projects, and teamwork: The good, the bad, and the ugly? *Journal of Marketing Education*, 21(2), 106-117.
- McCroskey, J. C., Gudykunst, W. B., & Nishida, T. (1985). Communication apprehension among

- Japanese students in native and second language. *Communication Research Reports*, 2(1), 11-15
- MEXT (2009) 中学校学習指導要領「外国語」英訳版（仮訳）[Junior high school Course of Study (foreign languages) English version (provisional translation)]. Retrieved from: http://www.mext.go.jp/component/english/_icsFiles/afiedfile/2011/03/17/1303755_013.pdf
- MEXT (2011) 高等学校学習指導要領英訳版（仮訳）外国語 [High school Course of Study, English version (provisional translation) foreign languages]. Retrieved from: http://www.mext.go.jp/a_menu/shotou/newcs/youryou/eiyaku/_icsFiles/afiedfile/2011/04/11/1298353_9.pdf
- MEXT (2013) 大学における教育内容等の改革状況について（平成25年度）[On the state of reforms to educational curriculum in universities (FY2013)]. Retrieved from: http://www.mext.go.jp/a_menu/koutou/daigaku/04052801/1361916.htm
- MEXT (2014) On integrated reforms in high school and university education and university entrance examination aimed at realizing a high school and university articulation system appropriate for a new era (report). Retrieved from: http://www.mext.go.jp/component/english/_icsFiles/afiedfile/2015/03/31/1353908_1.pdf
- Michaelson, L. K., & Black, R. H. (1994). Building learning teams: The key to harnessing the power of small groups in higher education. In S. Kadel & J. Keehner (Eds.), *Collaborative learning: A sourcebook for higher education* (Vol. 2, pp. 65-81). State College, PA: National Center for Teaching, Learning & Assessment.
- 水野邦夫, 田積徹, 炭谷将史, 多胡陽介 [Mizuno Kunio, Tazumi Tooru, Sumiya Masashi, and Tago Yousuke] (2008) 大学新入生の大学適応を促進する授業プログラムの検討（人間学部人間心理学科, 聖泉大学）[An examination of a teaching program for freshmen to facilitate adjustments in campus life] *聖泉論叢* [*Bulletin of Seisen University*], 15, 125-140.
- Mitchell, C. (2017). Language education pressures in Japanese high schools. *JALT Shiken*, 21(1), 1-11.
- Morgan, B. B., A. S. Glickman, E. A. Woodard, A. S. Blaiwes, and E. Salas. (1986). *Measurement of team behaviors in a navy environment*. Orlando, FL: Naval Training Systems Center.
- 森常人 [Mori Tokihito] (2016), 大学生のコミュニケーション能力向上のための基礎的研究：グループワークとプレゼンテーションの事例より [Basic research for improving the communication abilities of university students: From case study of group work and presentations], *関西外国語大学 研究論集* [*Kansai Gaidai College Journal of Inquiry and Research*], 104, 129-138.
- 中村和彦 [Nakamura Kazuhiko]. (2013). 大学1年春学期におけるラボラトリー方式の体験学習の効果—体験から学ぶ力の影響 [Experiential learning through laboratory method on university freshmen during their first semester: Effects of learning abilities from experience based on the ELAHE model]. *実験社会心理学研究* [*The Japanese Journal of Experimental Social Psychology*], 52(2), 137-151.
- 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.
- Nozaki, K. (1993). The Japanese student and the foreign teacher. In Wadden, P. (Ed.), *A handbook for teaching English at Japanese colleges and universities*. (pp. 27-33). New York: Oxford University Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory*, 3, 248-292.
- O'Neill, T. A., Goffin, R. D. & Gellatly, I. R. (2012). The knowledge, skill, and ability requirements for

- teamwork: Revisiting the Teamwork-KSA test's validity. *International Journal of Selection and Assessment*, 20, 36-52.
- Prichard, J. S., Bizo, L. A., & Stratford, R. J. (2006). The educational impact of team-skills training: Preparing students to work in groups. *British Journal of Educational Psychology*, 76(1), 119-140.
- Prichard, J. S., Stratford, R. J., & Bizo, L. A. (2006). Team-skills training enhances collaborative learning. *Learning and Instruction*, 16(3), 256-265.
- Rhee, J., Parent, D., & Basu, A. (2013). The influence of personality and ability on undergraduate teamwork and team performance. *SpringerPlus*, 2(1), 16.
- Riggs, M. L., Warka, J., Babasa, B., Betancourt, R. & Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job-related applications. *Educational and Psychological Measurement*, 54, 793-802.
- Rupp, M. J. (2016). Using focus groups for qualitative feedback on psychometric instruments. In G. Brooks, M. Porter, D. Fujimoto, & D. Tatsuki (Eds.), *The 2015 PanSIG Journal* (pp. 162-170). Tokyo, Japan: JALT.
- Rupp, M. J. (2017). Recommended modifications to the Kambara locus of control scale based on mixed methods analysis. *Proceedings of School of Agriculture, Tokai University*, 36, 23-36.
- Sakurai, Y. (2015). Implementing Cooperative Learning in EFL Reading Classes at a Japanese University. *言語と文化：愛知大学語学教育研究室紀要 [Language and Culture: The Bulletin of the Institute for Language Education, Aichi University]*, 32, 59-73.
- Schullery, N. M. & Gibson, M. K. (2001). Working in groups: Identification and treatment of students' perceived weaknesses. *Business Communication Quarterly*, 64(2), 9-30.
- 正保春彦 [Shobo Haruhiko]. (2012). グループワークの心理的効果についての一考察：構成的グループ・エンカウンターとインプロヴィゼーションの比較から [A study on the psychological effect of group work based on comparisons between structured group encounters and improvisation]. *茨城大学教育実践研究 [Studies in Teaching Strategies, Ibaraki University]*, 31, 279-291.
- Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20, 503-530
- Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, 25, 207-228
- 津村俊充 [Tsumura Toshimitsu]. (2010). グループワークトレーニング [Groupwork training]. *教育心理学年報 [The Annual Report of Educational Psychology in Japan]*, 49, 171-179.
- Van Duzer, E., & McMartin, F., (2000). Methods to improve the validity and sensitivity of a self/peer assessment instrument. *IEEE Transactions on Education*, 43, 153-158.
- Xethakis, L. (2017). Creating conditions for collaborative learning in the language classroom. In G. Brooks (Ed.), *The 2016 PanSIG Journal* (pp. 351-359). Tokyo, Japan: JALT.

Appendix – Groupwork Skills Questionnaire (GSQ) (Cumming et al., 2015)

Think about your usual contribution to group work. When answering the following questions, rate how frequently you have done the following when working in groups.

1 : Never 2 : Not very often 3 : Sometimes 4 : Quite often 5 : Always

When working in groups, I tend to...

- (1) provide emotional support to my group members.
- (2) remind the group how important it is to stick to schedules.
- (3) be sensitive to the feelings of other people.
- (4) construct strategies from ideas that have been raised.
- (5) show that I care about my group members.
- (6) clearly define the roles of each group member.
- (7) be open and supportive when communicating with others.
- (8) move the group's ideas forward towards a strategy.
- (9) be there for other group members when they need me.
- (10) evaluate how well the group is progressing towards agreed goals.

Interpersonal Skills (IS): Items 1, 3, 5, 7 and 9

Task Groupwork Skills (TGS): Items 2, 4, 6, 8 and 10

日本での第二言語獲得場面における日本語版グループワークスキル尺度の適用

セタキス・ラリー・ジョン

本研究では、第二言語獲得場面におけるグループワークスキル尺度（GSQ; Cumming, Woodcock, Cooley, Holland, & Burns, 2015）の適用について報告する。研究対象は日本における大学生である。研究の第一段階として、まずオリジナルの英語版尺度を邦訳した。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。西日本に所在する二つの大学から307名の回答を得、分析対象とした。調査項目の正規性が確認でき、二つの下位尺度の信頼性をクロンバック α 係数により確認した。確証的因子分析を用い、尺度開発者の仮説である二因子モデルへの適合度を分析したが、適合度は不良であった。続いて検証のため、GSQを構成する2つの下位尺度（対人スキル・チームワークスキル）ごとに確証的因子分析を行ったところ、十分ではないが尺度全体の分析に比べてモデルの適合度は改善した。これらの結果の示唆することや、このGSQの適用において考慮すべき改善点について考察を行った。