The Psychometrics of Affective, Cognitive and Social Factors Impacting Group-Based Learning Activities

Larry Xethakis

Summary

Introduction
Group-based learning approaches (GBLAs) are becoming increasingly prominent in Japanese language classrooms in response to educational policy reforms instituted by The Ministry of Education, Culture, Sports, Science and Technology (MEXT) as well as changes in pedagogical practices which emphasize communication and interaction. These approaches present a number of cognitive, affective and social challenges for students that differ from those in the traditional classroom. These challenges call for an expanded program of research to more fully examine their impact on GBLAs as well as aid in the identification of learners who may be put at risk by these changes in the educational environment. The research presented in this thesis aims to contribute to a psychometric foundation for such research.

The measurement studies reported in this thesis employed Confirmatory Factor Analysis (CFA) to investigate the psychometric properties of the following instruments: 1) the companion Interaction and Audience Anxiousness Scales (IAS-AAS; Leary, 1983a), concerned with aspects of social anxiety; 2) the Fear of Negative Evaluation Scale (FNE; Watson & Friend, 1969), concerned with social evaluative anxiety; 3) the Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994), concerned with aspects of metacognition; 4) the Feelings Towards Group Work Questionnaire FTGW (Cantwell & Andrews, 2002), concerned with learners’ dispositions towards group work; 5) the Groupwork Skills Questionnaire (GSQ; Cumming, Woodcock, Cooley, Holland, & Burns, 2015); concerned with the use of social skills to enhance the efficacy of group work.

Literature Review
Group-based learning approaches (GBLAs) are a family of pedagogical techniques that are finding their way into the Japanese educational context with increasing frequency. In the foreign language classroom, the use of pair-work and group-work activities is becoming more and more common. One reason for the increasing prevalence of GBLAs is the introduction of language teaching methodologies which emphasize meaningful communication and interaction as the basis for SLA, most prominently CLT, and the more recent associated approach of TBLT
(Leeming, 2012). These approaches take the development of communicative competence, that is the ability to use the language effectively and appropriately (Richards & Rogers, 2001; Richards & Schmidt, 2002), as their primary goal, and thus learner interaction in a social context has become central to language teaching in Japan. The need for interaction obviously entails the use of pairs or groups in the language classroom, and thereby the increased presence of GBLAs.

A further, and in some ways more fundamental, reason for the increasing emphasis given to GBLAs, are the progressive educational reforms that have been instituted by the MEXT over the past two decades. In the area of foreign language education, these reforms have shifted the emphasis from the acquisition of grammatical knowledge, or knowledge of language, to the acquisition of communication skills, or knowledge of how to use the language (e.g., MEXT 2011). These reforms have created a space for communicative language teaching methodologies to enter the classrooms, and with them have come GBLAs.

In addition to the reforms in language education, more recent reforms have brought GBLAs to prominence in other areas of the curriculum and at all levels of the educational system. The Second Basic Plan for the Promotion of Education (MEXT, 2013) promotes collaborative learning not only as a means of fostering learners’ academic abilities and communication skills, but also as an important component of a high-quality educational environment. Additional reforms (e.g., MEXT, 2012) have established active learning as a primary educational approach, which can provide learners with opportunities to develop the skills needed to function in the contemporary social and economic environment. These reforms envision the use of discussion and debates as a context for engagement with contemporary issues facing Japan, and this suggests that GBLAs play an important role in practically implementing active learning in the classroom. Finally, GBLAs have been closely linked with the introduction of information communication technology (ICT) into the Japanese educational system. Technology is not only seen as a key tool in encouraging student interest, but equally, as enabling students to teach and learn from each other (MEXT, 2016).

GBLAs undoubtedly have the potential to bring great benefits to learners. In the language classroom, GBLAs provide learners with more opportunities to speak as well as receive comprehensible input, and this aids their development of communicative competence (Lin, 2015). More generally, GBLAs can promote learner autonomy (Jones, 2007) and the development of critical thinking skills (Johnston, James, Lye & McDonald, 2000), as well as improved learning and overall academic performance (Johnson, Johnson & Smith, 1998).
Notwithstanding these benefits, it must be remembered that GBLAs also present a range of demands that are not present in the traditionally-structured and teacher-centered classroom. Working in groups often requires learners to deal with situations that are novel to them, that are ambiguous, and that involve working with individuals who are unfamiliar to them. Each of these factors has the potential to create feelings of unease in even well-adjusted learners. Furthermore, learners in a small group may be called upon to shift roles between teacher and learner, explain material and comprehend others’ explanations. This requires a significant degree of cognitive awareness. Finally, organizing groups to function effectively and managing interpersonal relationships in the group requires the deft use of social skills. For learners who are shy, or even worse, who already experience anxiety in social situations, who may not be cognitively prepared, or who lack the necessary social skills, working in groups may become an obstacle to learning rather than an aid. The increasing prominence of GBLAs in the Japanese educational context presents learners and educators with a number of new challenges. These new challenges would seem to call for a new trajectory of research to help establish the impact of affective, cognitive and social factors on the effectiveness of GBLAs, and moreover, to help identify and understand the situation of those learners who may not be prepared to function effectively in groups.

Rationale and Methodology
All research requires solid foundations, and this is particularly true when research is expanding into new areas or launching on a new trajectory. Solid foundations emerge only from evidence-based research, and this in turn, relies on accurate, valid and reliable observation and measurement for which there must also be evidence. Proper measurement is built on a clear understanding of the properties of the tools employed in making the observations, and such understanding comes only from empirical evidence for or against the accuracy, validity and reliability of these tools.

The research reported in this thesis makes an incremental advance in establishing the foundations for research on cognitive, affective and social factors that impact the efficacy of GBLAs, through the examination of the psychometric properties of a number of self-report instruments hypothesized to measure these constructs. The instruments included in this study were chosen on the basis of their prominence in the literature, and more importantly, a lack of evidence regarding the measurement properties of scores produced by each of these instruments in the Japanese university population. A total of six studies comprise the body of this thesis, and the studies examined the measurement properties of the following instruments:
1. The Interaction Anxiousness Scale (IAS; Leary, 1983a): this instrument is concerned with measuring subjective feelings of anxiety in contingent social encounters, e.g., conversations, speaking in small groups, etc.;

2. The Audience Anxiousness Scale (AAS; Leary, 1983a): this instrument is concerned with measuring subjective feelings of anxiety in non-contingent social encounters, e.g., giving a speech or presentation, etc.;

3. The Fear of Negative Evaluation Scale (FNE; Watson & Friend, 1969): this instrument assesses social evaluative anxiety, i.e., the degree to which respondents are concerned with others’ perceptions and evaluations of them in social encounters;

4. The Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994): this instrument is concerned with measuring respondents’ conscious knowledge of aspects of learning and the means they employ to regulate their learning;

5. The Feelings Towards Group Work Questionnaire (FTGW; Cantwell & Andrews, 2002): this instrument is concerned with measuring learners’ dispositions towards group work;

6. Groupwork Skills Questionnaire (GSQ; Cumming, Woodcock, Cooley, Holland, & Burns, 2015): this instrument is concerned with measuring respondents’ use of social skills to enhance the functioning and efficacy of group work.

The data sets utilized in these were gathered from 1,987 EFL students at four universities in Kumamoto prefecture. The central analytic process employed Confirmatory Factor Analysis (CFA), a form of Structural Equation Modeling, which is specifically concerned with the assessment of measurement models. Over the six studies reported in this thesis, a total of 35 models were tested. In order to determine the extent to which the structures specified in each of the models corresponded with the underlying structure of the scores in the data set, four goodness-of-fit indices—the root mean squared error of approximation (RMSEA), the standardized root mean square residual (SRMR), the Tucker-Lewis index (TLI), and the comparative fit index (CFI)—were employed, together with the $\chi^2$ (chi-square) test statistic, following the recommendations of Brown (2015). The cut-off values for each of the four indices applied in determining model fit were based on the recommendations of Hu & Bentler (1999).

**Results and Discussion**
The first study presented in this thesis examined the measurement properties of score generated by the companion IAS and AAS scales (Leary, 1983a). For this instrument a total of four measurement models were tested—three based on the correlated, two-factor structure originally hypothesized by Leary, and one based on the correlated, two-factor structure proposed by Okabayashi and Seiwa (1991) for use in the Japanese population. The results of the CFA performed on a model specifying the structure proposed by Okabayashi and Seiwa suggested that this model closely matched the structure underlying the scores generated by this instrument. The model specifying Leary’s originally hypothesized structure lacked a sufficient degree of fit with the scores. Further investigation of signs of localized strain in this particular model indicated the strong possibility that the reverse-scored items included in these scales formed a second factor underlying scores on this version of the instrument, and moreover, that this factor was due to a method-effect.

The FNE (Watson & Friend, 1969) was the subject of the second study reported in this thesis. Three versions of this instrument are prominent in the literature: the original 30-item FNE proposed by Watson and Friend; a shortened, 12-item version, the BFNE, proposed by Leary (1983b); and another short, 12-item version, the SFNE proposed for use in the Japanese population (Sasagawa, Kanai, Muranaka, Suzuki, Shimada & Sakano, 2004). All three of these measures were thought to have a single construct underlying their scores (e.g., Leary, 1983b; Oei, Kenna & Evans, 1991; Sasagawa et al., 2004), however, recently, this supposition has been brought into question (e.g., Duke, Krishnan, Faith & Storch, 2006; Nihei et al., 2016; Rodebaugh, Woods, Thissen, Heimberg & Chambless, 2004), with evidence that the reverse-scored items on each of these scales form a second factor underlying the structure of scores. Furthermore, and on the basis of this research, it has been suggested that scales comprised only of the respective straightforwardly-scored items on the BFNE and the SFNE could act as better measures of social evaluative anxiety than the full scale (Nihei et al., 2016; Weeks et al., 2005). For these reasons, the study reported in this thesis had two aims: first, to provide empirical evidence for or against the single-factor structure of the FNE, the BFNE and the SFNE; and, second, to examine the fit of single-factor models of the BFNE and SFNE comprised only of the respective straightforwardly-scored items on these two scales.

For the purpose of determining whether scores produced by the FNE, BFNE and SFNE exhibited a single-factor or two-factor structure, models specifying a single-factor structure, as well as a correlated, two-factor structure (with the reverse-scored items on each scale specified as indicators of the second factor) were tested for each instrument. The results of the CFAs carried out on the two measurement models for the FNE were inconclusive, with both models
displaying less than sufficient fit. However, the two-factor model exhibited a much greater degree of fit, thus suggesting that scores on the FNE lack a single-scale structure and thus should only be interpreted with great caution. For both the BFNE and the SFNE, two-factors models exhibited arguably adequate fit. One-factor models of each scale, conversely, exhibited poor fit. These results add to the evidence for a two-factor structure underlying scores on these instruments, and again suggest that scores on these scales be interpreted with care.

Single-factor models of the BFNE and SFNE comprised only of the respective straightforwardly-scored items on these two scales were specified and tested. However, both models surprisingly exhibited poor fit. An investigation into possible sources of poor fit in the model for the SFNE revealed the presence of method effects engendered by two items with closely related content. Two further one-factor models, each omitting one of these two items, were specified and tested. Both models exhibited exemplary fit with the scores in the data set, suggesting that a scale with one of these two items removed could serve as a valid measure of social evaluative anxiety.

The third study reported in this thesis examined the properties of the MAI (Schraw & Dennison, 1994). For this purpose models specifying four hypothesized structures for this instrument were tested: 1) the correlated, two-factor structure hypothesized by the original authors; 2) the hierarchal, three-factor structure proposed by Abe and Ida (2010); 3) the correlated, three-factor structure proposed by Niwa and Yamaji (2017); and, 4) the correlated, three-factor structure proposed by Teo and Lee (2012). Disappointingly, none of the four models exhibited sufficient fit and thus were rejected as plausible models. These results provide evidence against a two- or three-factor structure for scores generated by the MAI in the Japanese context.

The structure of scores on the MAI was also examined in the fourth study reported. This study examined the version of the MAI hypothesized by Yasuda (2016). This structure for this particular version of the MAI is more strongly underpinned by metacognitive theory than those versions for the instrument tested in the third study, and thus was examined in a separate study. The measurement model specified to test the correlated six-factor structure hypothesized by Yasuda displayed arguable good fit with the scores in the data set, and this result provides evidence for the utility of this scale in the Japanese context.

Study Five examined the structure of scores on the FTGW (Cantwell & Andrews, 2002). The results of the CFA on the correlated, three-factor model specified to test the structure originally hypothesized by Cantwell and Andrews showed that this model did not fit the structure of the scores. Subsequent to this result, diagnostic CFAs were carried out on single-
factor models for each of the instrument’s three subscales. Two of these models failed to fit the structure of the scores sufficiently and were rejected. One model however, that for the Distress in Group Work subscale, displayed meritorious fit. This result suggests that this subscale has the potential to form the core of an instrument aimed at assessing learners’ feelings of anxiety towards group work.

The final study reported in this thesis investigated the measurement properties of the GSQ (Cumming et al., 2015). A measurement model specifying the correlated, two-factor structure hypothesized by the original authors was tested, however, the results of the CFA failed to show sufficient fit between this model and scores generated by the instrument. Diagnostic models for both of the instrument’s subscales also performed poorly and were rejected. Examination of the item content suggested that the terminology employed in describing the interpersonal skills assessed by the instrument might be vague to Japanese students, and thus, that the provision of concrete examples of the actions signified by these terms might improve the measurement properties of this scale.

Conclusion
The results presented in this thesis provide positive evidence for the use of several of the instruments tested. First, Okabayashi and Seiwa’s (1991) version of the IAS-AAS exhibited adequate fit with the scores in the data set, and thus may prove useful in research on aspects of interactional anxiety, as well as more general social anxiety. Second, an arguably good degree of fit was shown between the structure of scores in the data set and Yasuda’s (2016) version of the MAI, providing an empirical basis for the use of this instrument in examining aspects of learners’ metacognition. Finally, a single-factor version of the SFNE, comprised of seven of the straightforwardly-scored items on this scale, displayed excellent fit with the scores, suggesting that, with further refinement and confirmation, this scale might become a useful tool for assessing learners’ social evaluative anxiety. It must be noted however, that all the findings in this study represent only one piece of evidence for or against the various structures tested, and thus further examination of these instruments is called for.

In addition to the positive evidence outlined above, a number of broader issues emerged in the course of this study, two of which will be mentioned here. The influence of reverse-scored items on the structure and scoring of self-report instruments (e.g., the PSWQ [Brown, 2003]; the SIAS [Rodebaugh et al., 2006]; the BFNE [Weeks et al., 2005]) has become an area of increasing interest in the literature. Several of the analyses conducted in the studies included in this thesis suggest the need for a broader re-evaluation of the structure, and thus the scoring
regimes, of instruments containing reverse-scored items. These results bear on the broader point that when reverse-scored items are included in an instrument, developers and practitioners should maintain an awareness of the possibility of method effects, which can detract from the validity of the scores generated by the instrument in question. The significant impact the presence of these items can have on scores would seem to provide further evidence for at least a preliminary re-evaluation of the structures, and scoring regimes, of instruments including reverse-scored items, particularly in the Japanese context, where such research is just beginning.

The tendency of measurement studies in the literature not to report information concerning the degree of non-normality found in the dataset or datasets upon which their analyses are based was also a point brought to light by the studies reported in this thesis. Given the importance of this information for the interpretation of the results of the EFAs or CFAs carried out in each of these studies, this would seem to be a case of pervasive under-reporting that should be rectified.

References

阿部真美子[Abe Mamiko] & 井田政則 [Ida Masanori]. (2010). 成人用メタ認知尺度の作成の試み：Metacognitive Awareness Inventory を用いて [An attempt to construct the adults’ metacognition scale: Based on Metacognitive Awareness Inventory]. 立正大学心理学部 [Journal of the Psychology Department of Rissho University], 1, 23-34.


the Faculty of Integrated Arts and Sciences, Hiroshima University, III, Studies in Information and the Behavioral Sciences, 15, 1-9.


