東海大学文理融合学部紀要　第２巻（2024年）25頁～38頁
Journal of the School of Humanities and Science, Tokai University, Vol.2. (2024) pp.25-38

The Psychometric Properties of the Short Fear of Negative Evaluation Scale in the Japanese SLA Context

by
Larry J. Xethakis* and Michael J. Rupp*

(Received: September 14, 2023, Accepted: December 12, 2023)

Abstract

This measurement study examined the construct validity of the Short Fear of Negative Evaluation Scale (Sasagawa et al., 2004), as well as variants of the scale proposed by Nihei et al. (2018), in a sample of Japanese university-level English learners (N = 411) from three universities in Western Japan. Confirmatory factor analysis of a unidimensional model of the Short Fear of Negative Evaluation Scale revealed poor fit between the model and the scores in the dataset. A two-factor model, with the four reverse-scored items specified to load on a separate factor displayed an adequate degree of fit, suggesting that the scale, as proposed by Sasagawa et al., is not unidimensional. A one-factor model of the eight straightforward items, proposed by Nihei et al. was also tested, and the model showed an inadequate degree of fit. Examination of areas of ill-fit in this model indicated the possibility of content overlap between two closely worded items (Items 2 and 5). Diagnostic models lacking these two items showed an excellent degree of fit, suggesting a scale without one of the items could serve as a potentially valid measure of fear of negative evaluation in the target population.

Key Words : Social anxiety, Fear of negative evaluation, English education, Group-work, Validity

1. Introduction

Group work and group-based learning approaches have become an integral part of language teaching. Communicative language teaching and task-based language teaching are two of the more commonly employed teaching methodologies in which interaction between learners form a significant part of instruction (Leeming, 2011). While group work can provide many benefits for learners, such as greater opportunities for communication and learning from peers (Johnson, et al., 1998), working in groups can also present challenges for those learners who experience unease when interacting with others (Cantwell & Andrews, 2002). Group-based learning approaches, which often employ novel activities and which force learners to work with new people and express themselves in the L2, can be much more anxiety inducing than the more traditional, teacher-centered classroom many learners are accustomed to. In particular, the ambiguity and uncertainty associated with these situations can be difficult for learners to deal with and thus provoke feelings of social anxiety in learners (Zhou, 2016). While research into the impact of social anxiety on language learning has begun, King and Smith (2017) note that there is a need for more research in this area. For such research to proceed, there is a need for research into the reliability and validity of instruments to measure social anxiety when employed in the language learning context (Xethakis, 2020). This study represents an attempt to address this need by examining the reliability and validity of the Japanese version of the Short Fear of Negative Evaluation scale (SFNE; Sasagawa et al., 2004).

*Associate Professor, Department of Community and Social Studies. Tokai University

2. Literature Review

The psychological construct of fear of negative evaluation can be characterized as circumscribing a “broad social evaluative anxiety,” (Collins, Westra, Dozois & Stewart, 2005). More specifically, it is involved with a feeling of unease, or in the extreme, a sense of overwhelming apprehension, at being evaluated in an unfavorable or critical manner in social situations (Weeks et al., 2005). A significant feature of social anxiety is an individual’s sense of apprehension with respect to their own ability to manage how they are perceived and evaluated by others, and thus fear of negative evaluation and the broader construct of social anxiety are commonly regarded as closely related constructs (Schlenker & Leary, 1982). For this reason, fear of negative evaluation can be considered as one of the “central cognitive aspects of social anxiety,” (Leary, 1991, p. 166).

The Fear of Negative Evaluation scale (FNE; Watson & Friend, 1969) is a widely-employed instrument in assessing social evaluative anxiety (Rodebaugh et al., 2004). This measure is postulated to assess individual differences in the degree to which respondents are concerned about how they are perceived and evaluated by others (Leary, 1991). The measure was originally developed by Watson & Friend (1969) and comprises 30 items, scored using a dichotomous, true-false scale. The items are concerned with various facets of unfavorable evaluation in social situations, for example*, I am afraid that others will not approve of me* (Item 13), or *I often worry that I will say or do the wrong things* (Item 25). Seventeen of the 30 items comprising the instrument are straightforward items, which describe cognitions of worry or unease, while the remaining 13 are reverse-scored, with Watson & Friend (1969) hypothesizing these to express the opposite of fear of negative evaluation, that is a lack of unease or discomfort at being evaluated by others.

The instrument was initially developed in a sample of 205 university students, and displayed a high degree of internal consistency—with values of .94 (using the Kuder-Richardson Formula 20) in the original sample, and .96 in a second sample of 154 students (Watson & Friend, 1969). The scale also displayed a good degree of convergent validity. The FNE was found to correlate significantly with measures of manifest anxiety ($r = .60$), social evaluative anxiety ($r = .47$), and need for social approval ($r = .60$).

The FNE has been adapted for use in the Japanese population by Ishikawa, Sasaki and Fukui (1992), using scores from a sample of 332 individuals, of which 250 were university students, 50 were individuals from the community, and the remaining 32 individuals were in clinical treatment. Participants responded to each item on a dichotomous true-false scale. Upon analysis, scores on each of the items were found to correlate significantly with scores on the scale as a whole. The structure of the scores was analyzed using Principal Component Analysis. As a result of this analysis, the first factor extracted explained 38% of the total variance, while the second factor extracted explained less than 7% of the variance. From this the authors concluded that the Japanese version of the FNE was primarily unidimensional in nature.

The instrument was also found to correlate highly with Japanese versions of the Manifest Anxiety Scale ($r = .67$), and the State-Trait Anxiety Inventory ($r = .62$), as well as to discriminate sufficiently between individuals with high and low social evaluative anxiety. This version of the FNE has been employed extensively in the Japanese population, with more than 20 studies that have made use of this measure published as of 2023.

In addition to its popularity as a measure of social evaluative anxiety, Ishikawa et al.'s (1992) version of the FNE has also served as the basis for a shortened version of the FNE. This measure, known as the Short Fear of Negative Evaluation scale (SFNE; Sasagawa et al., 2004), is one of the most commonly used instruments for assessing social anxiety and its cognitive aspects in particular in Japan (Maeda et al., 2017). The SFNE has been employed in nearly 50 studies examining social anxiety, including two dozen studies published in the English literature concerned with this field of research. This shorter version of the FNE was developed by Sasagawa and her colleagues in response to shortcomings in the 30-item Japanese FNE, which while seen as an adequate instrument, was thought to be somewhat long (Sasagawa et al., 2004). Moreover, by changing the scoring regime from a dichotomous true-false scale, used on most versions of the FNE, to a 5-point Likert scale, Sasagawa et al. hoped to improve the discriminative power of the instrument.

The SFNE was developed on the basis of scores from 389 university students on the thirty items of the Japanese version of the FNE (Ishikawa et al., 1992). Participants were asked to respond on the degree to which each item was characteristic of them, with 1 semantically anchored to *not at all characteristic of me*, and 5 to *extremely characteristic of me*. From the 30 items on the original FNE, the 12 highest-loading items from the factor pattern in the study were chosen to form the SFNE. Since its development, the SFNE has shown a high degree of reliability, with reported values for Cronbach's alpha ranging from .85 (Moriya & Tanno, 2011) to .93 (Curtis et al., 2017), and with the majority of studies reporting values of .90 or greater (e.g., Okajima et al., 2009; Maeda et al., 2017). Scores on the SFNE have also been shown to correlate with scores on other measures commonly used in the assessment of social anxiety, such as the SPS ($r = .57$, $p < .05$; Okajima et al, 2009); the SIAS ($r = .62$, $p <. 01$; Essau et al., 2012); the SAD ($r = .48$, $p < .01$; Shirotsuki et al., 2012); and the PSWQ ($r = .58$, $p < .05$; Maeda et al., 2015), among others.

Initially, the factor structure of the FNE, in both its English and Japanese versions, was assumed to be unidimensional. This was also thought to be the case for the SFNE as well. However, this supposition has recently been brought into question. Rodebaugh et al. (2004) performed the first analyses on the FNE for the purpose of determining if scores on the instrument possessed the property of unidimensionality. For this purpose, Rodebaugh and colleagues tested two similar measurement models for the FNE. The first model was a one-factor model, with all items specified to load onto this single factor. The second measurement model was a two-factor model, with the straightforward (straight-scored) items hypothesized to load onto one factor and the reverse-scored items to load onto a second factor, with these two factors allowed to correlate. The factor structure of the second model was based on prior research (e.g., Brown, 2003) showing that reverse-scored items which are included in instruments exhibit a tendency to form separate factors.

On the basis of results from confirmatory factor analysis (CFA) on the two models, Rodebaugh and his colleagues concluded that the two-factor model of the FNE better described the underlying structure of the scores in the respective datasets. They suggested two possibilities for the better fit of the two-factor model. First, the factor comprising the reverse-worded items might be the result of a method effect brought about by the wording of the reverse-scored items. This was accepted as the most parsimonious explanation for the two-factor structure. However, they also stated that given the methodology of the study, it was not possible to rule out the second possibility, that the reverse-worded items comprised a substantial construct rather than a method effect.

After determining that the two-factor model fit the data better, the discriminative power of the straightforward and reverse-scored factors on both instruments was investigated, as well their concurrent validity, that is, their ability to predict scores on other social anxiety measures. It was found that the items comprising the straightforward factor were found to better discriminate across a range of degrees of fear of negative evaluation. It was supposed that this was because the straightforward items were less confusing than the reverse-scored items. Furthermore, the straightforward items also performed better in terms of concurrent validity, being significantly more predictive of scores on other self-report social anxiety measures than were the reverse-scored items. The results led Rodebaugh and colleagues to suggest that the straightforward items might represent a better measure of social anxiety than the full instrument, which included the reverse-scored items. However, they noted that these conclusions needed to be examined in further research.

Nihei et al. (2018) undertook an examination of the structure of scores produced by the SFNE (Sasagawa et al., 2004), on the basis of the findings of Rodebaugh et al. (2004). The analysis was based on responses from a community sample of 500 participants between the ages of 20 and 69, as well as a smaller sample of 82 university students. The survey was presented to participants online, and scores were recorded using a 5-point Likert scale. Participants responded with the degree to which the items were characteristic of them, with 1 being semantically anchored to, *not at all*, and 5 to *extremely characteristic of me*. Initially, exploratory factor analysis (EFA) with Promax rotation was carried out and evidence was found for a two-factor solution, with the first factor comprised of the eight straightforward items on the SFNE, and the second comprised of the four reverse-scored items. Reliability estimates (Cronbach's alpha) for scores on factors comprised of the straightforward items (hereinafter, SFNE-S) and the reverse-scored items (SFNE-R) were calculated, with values of .94 and .86 being derived, respectively.

Subsequent to the EFA, two measurement models were specified in order to clarify the dimensionality of the scores on the SFNE using CFA. The first model was a one-factor model of the SFNE, with all 12 items specified to load on the single factor, while the second model comprised two correlated factors, with the eight straightforward items specified to load on one factor, and the reverse-worded items specified to load on a separate factor. The results of CFA on the one-factor model indicated unequivocally poor fit, which suggested that scores on the SFNE did not possess a simple unidimensional structure, consistent with the results of the EFA.

The results for the two-factor model of the SFNE, were more ambiguous, with some indices suggesting adequate fit, and some unsatisfactory fit. Overall, while the results of the CFA performed by Nihei et al. (2018) did not present conclusive evidence that scores produced by the SFNE are comprised of two dimensions, the results do seem to suggest that the SFNE, in a manner similar to the FNE, also possesses a two-factor structure.

In addition to examining the structure of the SFNE, Nihei et al. (2018) also investigated the convergent validity of the SFNE, the SFNE-S and the SFNE-R. Scores on both the SFNE and the SFNE-S were found to correlate significantly with scores on three other instruments used to assess aspects of social anxiety, the LSAS, the Fear of Positive Evaluation scale and the Generalized Anxiety Disorder Questionnaire. Scores on the SFNE-R displayed a degree of correlation with scores on the LSAS and the Generalized Anxiety Disorder Questionnaire, but to a much lesser degree than either the SFNE or the SFNE-S. Scores on the SFNE-R did not correlate significantly with scores on the Fear of Positive Evaluation scale. On the basis of their results, Nihei et al. (2018) suggested that the SFNE-S, that is, the scale comprised of the eight straightforward items from the SFNE, be used in future research on social anxiety, rather than the full SFNE (Sasagawa et al., 2004).

Given the increasing emphasis placed on group work in language classroom (Leeming, 2011) and the negative impact social anxiety can have on the learners' willingness to engage in group work (e.g., Cantwell & Andrews, 2002; Zhou, 2016) there is a need for a valid and reliable instrument to assist both researchers and practitioners in the assessment of learners' self-perceptions of social anxiety. As outlined above, the construct of fear of negative evaluation can be seen as central to conceptions of social anxiety, and thus the accurate measurement of this construct is of importance for the progress of research in this area.

While the results of Nihei et al. (2018) do provide evidence suggesting that the SFNE also has a two-factor structure, the results of the CFAs carried out were ambiguous in relation to the two-factor model, and thus the structure of scores produced by the SFNE remains indeterminate. In view of the wide-spread use of the SFNE (Sasagawa et al., 2004) in the Japanese context, and the fact that the structure of scores produced by this measure, and thus importantly the validity of these scores, has been brought into question, confirmation of the structure of scores produced by this instrument is called for. This study seeks to address this need and extend the findings of Nihei et al. (2018), by examining the factor structure of the SFNE, as an incremental step to placing the use of these instruments on a stronger, evidence-based, footing.

This study aims to clarify the ambiguity by testing both one- and two-factor models of the SFNE with a new dataset in order to determine if a two-factor model of the SFNE adequately reflects the underlying structure of scores produced by the measure. If this were found to be the case, then in accordance with the findings of Nihei et al. (2018), the SFNE-S (that is, the subscale comprising only the straightforward items from the SFNE) would appear to be a more appropriate measure than the full SFNE for evaluating social evaluative anxiety in the Japanese context, and for this reason, this study also aims to examine the structural validity of the SFNE-S as an independent measure for assessing social evaluative anxiety in the Japanese population.

3. Methodology

3.1 Participants

The dataset employed in this study was gathered from students enrolled in English classes at three universities, one public and two private, located in Western Japan. A total of 411 responses were collected. Upon initial inspection, five responses were found to have missing data, and were therefore removed from the dataset. As no clearly discernable pattern was found in the missing responses (this was determined through visual inspection by the authors), the removal of these responses was not considered to have an effect on the overall characteristics of the dataset. The data from the remaining 406 responses forms the basis for the analyses presented below.

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

3.2 Instrument

The SFNE scale (Sasagawa et al., 2004) is the shortened version of the Japanese adaptation of the FNE (Ishikawa et al., 1992). It comprises twelve items from the Japanese version of the FNE (See Appendix) It also employs a 5-point Likert scale to measure the degree to which respondents feel each statement is characteristic of them, with 1 semantically anchored to *not at all*, and 5 to *extremely*. Eight of the items comprising the scale are scored straightforward and four are reverse-scored.

3.3 Analytical Procedures

SPSS v21 was used to calculate descriptive statistics (means, standard deviations, and the degree of skew and kurtosis) for each item comprising the SFNE. The univariate normality of the scores was determined by dividing the skew and kurtosis values of each item by their respective standard errors to calculate critical ratios, which are compared to a strict criterion of 2.0 and a more relaxed criterion of 3.0. Mardia's Coefficient was employed to determine the degree of multivariate normality. Cronbach's alpha with 95% confidence intervals (Fan & Thompson, 2001) was calculated for the SFNE and the variations tested in this study. Following the recommendations of Nunnally and Bernstein (1994), a value of .70 or greater for reliability was adopted. Finally, AMOS v21 was used to conduct CFA on models of the SFNE and its variations. A total of five models were tested, outlined in the Results section below. To determine the degree of model fit, the chi-square (χ^2) was used in conjunction with 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). The combination of a range of fit indices together with the chi-square is one means to overcome the latter's tendency to over-reject models. The values recommended by Hu and Bentler (1999) as indicating a good degree of fit for the fit indices (TLI and CFI >.95; RMSEA <.06;

SRMSR <.08), were used to evaluate the fit of the model of the SFNE and each of its variants.

4. Results

4.1 Descriptive Statistics, Skew and Kurtosis

Table 1, below, shows the descriptive statistics for the 12 items comprising the SFNE. The range of the means extends from 2.04 (Item 5) to 3.76 (Item 4). For the standard deviations, the largest value was found to be that for Item 10, at 1.269, while the smallest standard deviation was for Item 5, at 1.100.

Table 1: Item Means, Standard Deviation, Skew and Kurtosis for the Items on the SFNE

Item	M	SD	Skew	Std. Error	Kurtosis	Std. Error
1	3.06	1.152	0.006	0.121	-0.754	0.242
2	2.14	1.102	0.794	0.121	-0.100	0.242
3	3.39	1.224	-0.431	0.121	-0.708	0.242
4	3.76	1.132	-0.766	0.121	-0.165	0.242
5	2.04	1.100	0.932	0.121	0.103	0.242
6	2.56	1.140	0.351	0.121	-0.637	0.242
7	2.64	1.135	0.323	0.121	-0.636	0.242
8	2.12	1.140	0.822	0.121	-0.203	0.242
9	3.59	1.195	-0.630	0.121	-0.535	0.242
10	2.67	1.269	0.263	0.121	-1.002	0.242
11	3.61	1.252	-0.663	0.121	-0.537	0.242
12	2.40	1.163	0.558	0.121	-0.574	0.242

The calculated critical ratios for the skew and kurtosis of each item are shown in Table 2. In terms of skew, the performance of the majority of the items was found to be less than ideal, with only one of the items possessing a critical ratio less than the strict threshold. The ratios for three of the 12 items (25%) fell between 2.0 and 3.0, while those for the remaining eight items surpassed the more relaxed threshold. The critical ratios for kurtosis exhibited better results, with values for 33.3% of the items falling below the 2.0 threshold. Those for 50% of the items were between the strict and more relaxed value, and only 2 of the items greater than the relaxed value. Overall, the results for the respective critical ratios for the skew and kurtosis of each of the items indicates a degree of univariate non-normality in the scores that make up the dataset. However, it should be noted that Sasagawa et al. (2004) did not report the degree of non-normality found in their dataset, and thus it is difficult to tell if the degree of non-normality described above is particular to the dataset used in this study or a more general characteristic of the instrument when used in the Japanese population. If a similar degree of non-normality was inherent in the data from the original study, then it might be the case that the non-normality found in the scores of the dataset employed in this study is not simply particular to this study, but a more invariant characteristic of the instrument itself.

Table 2: Critical Ratios for Skew and Kurtosis for each Item on the SFNE

Item	Skewness	Kurtosis
	Calculated Values	Calculated Values
1	0.05	**3.12
2	**6.56	0.41
3	**3.56	*2.93
4	**6.33	0.68
5	**7.70	0.43
6	*2.90	*2.63
7	*2.67	*2.63
8	**6.79	0.84
9	**5.21	*2.21
10	*2.17	**4.14
11	**5.48	*2.22
12	**4.61	*2.37

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

4.2 Reliability Estimates

Table 3 displays Cronbach's alpha reliability estimates, with 95% confidence intervals, for the three variations of the SFNE examined in this study. The values for the SFNE and its variants exceed the recommended value of .70 (Nunnally & Bernstein,

Table 3: Reliability Estimates, Confidence Intervals for Alpha (95%), Scale Means, and Scale Standard Deviations for Scores on the SFNE

Subscale	Cronbach's alpha	95% Confidence Intervals		Scale Mean	SD for Scale
		Lower Bound	Upper Bound		
SFNE	.889	.872	.904	33.97	9.401
SFNE-S	.901	.885	.915	19.62	7.073
SFNE-R	.813	.781	.841	14.34	3.847

1994), with the lower bounds of the 95% confidence intervals for all of the scales exceeding this threshold. These results suggest that all of the scales examined in this study possess sufficient reliability, at least in regards to this dataset. The value for the SFNE is similar to those reported in the literature; however, it is interesting to note that the reliability estimates for the SFNE-S and SFNE-R listed in Table 5.3 are lower than those reported by Nihei et al. (2018; .94 and .86, respectively).

4.3 Confirmatory Factor Analysis

A total of five measurement models were tested using CFA in this study. The results from each CFA are reported below in their respective sections. For the SFNE, two models were tested—a one-factor model, and a two-factor model. Following this, the results for the CFA carried out on a model of the SFNE-S are reported. The final section comprises the results from CFAs on two rival models of the SFNE-S. Table 4 presents the values for the goodness-of-fit indicators and χ^2 for each of the models.

Table 4: Comparison of Goodness-of-fit Indicators for SFNE and Variants

	SFNE One-factor model	SFNE Two-factor model	SFNE-S One-factor model	SFNE-S Rival Model 1	SFNE-S Rival Model 2
TLI	.729	.928	.920	.984	.991
CFI	.778	.942	.943	.989	.994
RMSEA	.153	.079	.107	.050	.037
SRMR	.1100	.0513	.0441	.0249	.0211
χ^2	564.61	185.43	113.04	28.29	21.87
p	.000	.000	.000	.013	.081

TLI: Tucker-Lewis index; CFI: Comparative fit index; RMSEA: root mean squared error of approximation; SRMR: standardized root mean squared residual; $\chi 2$: Chi-square test statistic.

4.3.1 One-Factor Model for the SFNE

The first measurement model specified for the SNFE was derived from the structure hypothesized by Sasagawa et al. (2004). In their study, all twelve items were found to load onto a single factor. For that reason, this model possessed a single factor with all twelve of the items comprising the SFNE specified to load onto that factor. The model possessed 78 distinct sample moments, 24 distinct parameters to be estimated, with 54 degrees of freedom, and thus was overidentified.

The value of Mardia's coefficient for this model was 24.226, indicating a degree of multivariate non-normality in the scores for this model. The χ^2 value for this model was 564.61 with a probability level of .000. The results from the calculations of the fit indices were as follows (Hu and Bentler's [1999] cut-off values are given in parentheses): TLI .729 (>.95), CFI .778 (>.95); RMSEA .153 (<.06); SRMR .1100 (<.08). The combination of these values strongly suggests that the single factor model of the SFNE does not fit the underlying structure of the scores in the dataset. This result was not unexpected, and would seem to confirm the finding of Nihei et al (2018), who also found that the one-factor model did not exhibit a sufficient degree of fit.

4.3.2 Two-Factor Model for the SFNE

A two-factor model for the SFNE was tested next. This model is similar to that employed by Nihei et al. (2018) in their study of the factor structure of the SFNE. The two-factors in this model were allowed to correlate. Items 1, 2, 5, 6, 7, 8, 10 and 12 (the eight straightforward items on the SFNE) were specified to load on the first factor, while Items 3, 4, 9 and 11 (the four reverse-scored items) were specified to load on the second factor. This model also met the criterion for overidentification, with 78 distinct sample moments, 25 distinct parameters to be estimated, and therefore 53 degrees of freedom.

The results of the goodness-of-fit indices for this model were as follows (Hu and Bentler's [1999] cut-off values in parentheses): TLI .928 (>.95), CFI .942 (>.95); RMSEA .079 (<.06), with the 90% confidence interval ranging between .066 and .091; SRMSR .0513 (<.08). The χ^2 value was 185.43 with a probability level of .000. Mardia's coefficient for this model (24.226) was the same as that for the single-factor model above, which is as expected as both models comprise the same items. This value indicates multivariate non-normality in the data. These results, taken in conjunction, suggest that this two-factor model of the SFNE possesses an arguably sufficient degree of fit with the structure of the scores in the dataset. Departures on the TLI and CFA are small, the SRMSR result is good, and values of up to .08 on the RMSEA can be considered acceptable. While the results from the CFA conducted by Nihei et al. (2018) on a two-factor model of the SFNE did not conclusively indicate that such a model possessed a sufficient degree of fit, these results from the CFA undertaken in this study would appear to provide stronger evidence for the fit of this hypothesized model.

4.3.3 One-Factor Model for the SFNE-S

In addition to presenting evidence of a two-factor structure underlying the SFNE, Nihei et al. (2018) also examined the concurrent validity of scales comprised of the SNFE's straightforward items (SFNE-S) and its reverse-scored items (SFNE-R). Nihei et al. found that the SFNE-S was a better predictor of scores on other social anxiety instruments, and as a result, recommended the use of the SFNE-S over the full SFNE. As mentioned above, the present study seeks to examine the structural validity of this model. For this purpose, a measurement model was developed on the basis of the SFNE-S, as proposed by Nihei et al. (2018). This model was a one-factor model with the eight straightforward items comprising the SNFE (Items 1, 2, 5, 6, 7, 8, 10 and 12) loading on this single factor. The model possessed 36 distinct sample moments and 16 distinct parameters to be estimated, with 20 degrees of freedom, and thus was overidentified.

Mardia's coefficient for this model, 26.537, indicated a degree of multivariate non-normality in the dataset for this model. The χ^2 value for this model was 113.04 with a probability level of .000. The results from the calculations of the fit indices (with Hu and Bentler's [1999] cut-offs in parentheses) were as follows: TLI .920 (>.95), CFI .943 (>.95); RMSEA .107 (<.06); SRMR .0441 (<.08). Taken together, the four goodness-of-fit indices, as well as the χ^2 value, strongly indicate that this one-factor model for the SFNE-S does not fit the data to a sufficient degree and thus should be rejected.

The SFNE-S has been reported to have greater utility than the full SFNE by Nihei et al. (2018), due to the effect of the reverse-scored items on the structure of that instrument, and for that reason, it was decided to undertake an examination of the possible sources of the poor fit exhibited by the SNFE-S in the CFA described above. When investigating possible sources of mis-fit in measurement models, it is recommended to examine both the modification indices for the model in question, as well as its standardized residuals (e.g., Brown, 2015; Kline, 2011). However, it is also recommended that any changes to the model suggested by such an examination (also known as a specification search) also be supported by a theoretical rationale in order to avoid capitalizing on the chance variations occurring in the dataset under examination (Kline, 2011). With this caveat in mind, the modification indices and the standardized residuals for the model specified for the SFNE-S are presented below.

Table 5 shows the covariance modification indices for this model. In this table, only those modification indices whose value is 10 or greater are shown. The reason for this is that altering the model to account for relationships whose index is less than 10 often has little or no effect on the overall fit of the model (Byrne, 2016). By definition, the error terms for each item in a measurement model are assumed to vary independently. A covariance modification index with a value greater than 10 suggests the fit of the model would be improved if the error terms for the two items listed were allowed to correlate in the model. This suggested correlation between the error terms implies that the two indicators (items) in question are being affected by something that is not explicitly specified in the measurement model (Kline, 2011).

Table 5: Covariance Modification Indices for the SFNE-S Measurement Model

Covariance			Modification Index
e2	< >	e5	75.512
e1	< >	e10	10.065

As can be seen in Table 5, only two modification indices possessed values greater than 10, however, only one of these is markedly larger than the cut-off. This is the modification index for items 2 and 5, with a value of 72.512. The magnitude of this value suggests that there may be a substantial influence on responses to these two items that is not accounted for by the model as presently specified.

In addition to an inspection of the modification indices, examination of the model's standardized residuals is recommended when seeking possible sources of mis-fit (Brown, 2015). These residuals show the amount of covariance between items that is not explained by the relationships as specified in the model. Positive values indicate that the model is underestimating the relationship between the two items in question, while negative values suggest that the model overestimates the relationship between the two items. For standardized residuals, an absolute value greater than 1.96 is commonly employed as a threshold for determining which residuals may be problematic in a given model. This value corresponds to a statistically significant z score (p =.05), and this in turn implies that the unaccounted-for covariance is more than likely due to factors other than chance associations in the dataset (Brown, 2015). Byrne (2016) further clarifies this point by citing Jöreskog and Sörbom (1993), who note that values greater than 2.58 should be considered large and thus possibly problematic.

These values for the model of the SFNE-S are given in Table 6. Upon inspection of the table, only one value, that of the residual between Item 2 and Item 5, exceeds Brown's cut-off. Furthermore, this value (3.46) surpasses Jöreskog and Sörbom's (1993) threshold value as well. The sign of this residual is positive indicating that the model is underestimating the relationship between these two items. This result corresponds to the evidence from the inspection of the modification indices outlined above, and taken together, strongly suggests that there is something not accounted for in the measurement model that is influencing responses to these two items; with this something being some form of latent.

Table 6: Standardized Residuals for the SFNE-S Measurement Model

Item No.	12	10	8	7	6	5	2	1
12	0							
10	0.52	0						
8	0.46	-0.05	0					
7	-0.24	-0.10	0.10	0				
6	0.48	0.16	-0.43	0.31	0			
5	-0.75	-0.76	0.76	-0.40	-0.98	0		
2	-0.88	-0.79	-0.25	-0.07	-0.07	3.46	0	
1	0.25	1.31	-0.73	0.47	0.39	-1.16	-1.09	0

On the basis of this evidence, a further measurement model was specified to test the hypothesis that the relationship between Items 2 and 5 was a source of the poor fit exhibited by the model of the SFNE-S. Kline (2011) suggests the specification of an error covariance between the two items as a means to test a hypothesis such as this. For this reason, a measurement model which was identical to this model, with the exception of the addition of an error covariance specified between Items 2 and 5, was specified and tested.

This model had 36 distinct sample moments and 16 distinct parameters to be estimated, with 20 degrees of freedom, and thus was determined to be overidentified. The value for Mardia's coefficient for this model (21.136) indicated that a degree of multivariate non-normality was present in the scores from the dataset. The values for the goodness-of-fit indices were as follows: TLI .986 (>.95), CFI .990 (>.95); RMSEA .045 (<.06); SRMR .0246 (<.08). These values satisfy the cut-off put forth by Hu and Bentler (1999), and thus, indicate a more than adequate degree of fit for this model. The χ^2 value for this model was statistically non-significant (at the $p < .01$ level), $\chi^2 = 34.78$, $p = .015$, which suggests that the model does not differ from the structure underlying the scores, and therefore should be accepted. Of course, while this model fits well, it is a diagnostic model with the common-factor specification for unidimensionality sacrificed through the allowing of two error terms to covary.

The results from the CFA on this model would seem to confirm the hypothesis that the covariance between Items 2 and 5, which is unaccounted for by the model of the SFNE-S as initially specified, is a significant source of the poor fit displayed by this model. In order to limit the possibility that this was merely the result of chance associations in the dataset, as well as to attempt to determine a possible source of the covariance, it was decided to examine the content of the two items. Upon inspection, it was determined that covariance of these items was likely due to a method effect engendered from what is termed *content overlap* (Brown, 2003). This effect stems from the inclusion of similarly worded items, or of items which cover similar aspects of the underlying construct. This similarity in item content results in a greater degree of covariance between the two items than would otherwise be expected. If these two items (Item 2, *I am frequently afraid of other people noting my shortcomings*; Item 5, *I am afraid that people will find fault with me*) display such a high degree of similarity, then the principle of parsimony would suggest that one of them should be removed from the instrument. This would have the result of nullifying the method effect, and thus, hopefully, improving the overall fit of the SFNE-S scale. In order to test this hypothesis—that the removal of either Item 2 or Item 5 would improve the fit of the model of the SFNE-S—two further measurement models were specified, and each is described in its respective section below.

4.3.4 First Rival Model for the SFNE-S

The first rival model tested comprised seven items from the SFNE (Items 1, 5, 6, 7, 8, 10 and 12; Item 2 was removed from the model) all specified to load on one factor, which was hypothesized to represent the construct of fear of negative evaluation. This model met the criteria for overidentification, with 28 distinct sample moments, 14 distinct parameters to be estimated, and 14 degrees of freedom.

The model exhibited a degree of fit similar to that displayed by the model described immediately above (that with an error covariance specified between Item 2 and Item 5). The χ^2 value for this model was not statistically significant (at $p < .01$), $\chi^2 = 28.29$, $p = .013$, indicating that the model does in fact fit the scores to a sufficient degree. Mardia's coefficient, however, remained above the cut-off value (5.0), at 20.862 indicating some multivariate non-normality. The four goodness-of-fit indices for this model all satisfied the cut-offs recommended by Hu and Bentler (1999; given in parentheses), and were as follows: TLI .984 (>.95), CFI .989 (>.95); RMSEA .050 (<.06), with the range of the 90% confidence interval calculated to be from .022 to .077; SRMR .0249 (<.08).

In accordance with the evidence provided by the values for the indices, together with the non-significant result for the χ^2 test statistic, it was determined that this model displayed a meritorious degree of fit with the scores in the dataset for this study.

4.3.5 Second Rival Model for the SFNE-S

As the first rival model tested, removed Item 2 from the SFNE-S, the next model omitted Item 5. This model comprised a single factor, with seven of the eight straightforward items from the SFNE-S (Items 1, 2, 6, 7, 8, 10 and 12) specified to load on this factor. The model possessed 28 distinct sample moments and 14 distinct parameters to be estimated, with a resultant 14 degrees of freedom, which meant that the model was overidentified. As with all of the previous models examined in this study, the value for Mardia's coefficient (16.540) indicated a degree of multivariate non-normality in the scores for this model, although this was the lowest value for all the models in the study. The results for the fit indices were as follows (with Hu and Bentler's [1999] cut-offs given in parentheses): TLI .991 (>.95), CFI .994 (>.95); RMSEA .037 (<.06), with the range of the 90% confidence interval calculated to be from .000 to .066; SRMR .0211 (<.08). These values indicate that Model 12 also exhibits a meritorious degree of fit. Finally, the value for the χ^2 test statistic was statistically non-significant, $\chi^2 = 21.87$, $p = .081$, once again indicating that the relationships specified in the model sufficiently reproduce the covariance structure of the scores in the dataset.

Both of the two rival models exhibited an extremely high degree of fit, and on the basis of the values, the second rival model might be considered to be the better of the two, however, the values of most indices differ by only a small degree, and this may be more due to small variations related to chance, rather than a significantly better degree of fit for the second model. Moreover, evidence for the validity of any model comes out of a cumulative process, and thus the results of any one study can neither fully confirm nor reject any model beyond doubt. Therefore, the determination of which version of the SFNE-S is a more viable instrument going forward would seem to require that both models be tested against a new dataset in future research.

5. Discussion

This study examined the structure of scores generated by the SFNE (Sasagawa et al., 2004). This instrument is extensively employed in the measurement of social evaluative anxiety, as well as more general social anxiety. However, the dimensionality of the instrument, and thus also the valid interpretation of scores generated by it, has recently been brought into question (Nihei et al., 2018). The aim of this paper was to investigate the dimensionality of scores produced by this instrument, in order to provide further evidence for or against the factor structures suggested by previous research (e.g., Nihei et al., 2018; Rodebaugh et al., 2004). The determination of the structural validity of this instrument is an important initial step towards putting the further use of the SFNE on an evidence-based footing. If a positive determination were made, the SFNE could serve as a valuable tool in the investigation of the role of social anxiety in group work and, even more importantly, as means to assist educators in recognizing those learners who may be put at risk by the increasing emphasis on the use of group work in Japanese classrooms.

Investigation of the SFNE resulted in evidence against a simple unidimensional structure underlying scores on this instrument. As the results of the CFA showed, a single factor model of this instrument exhibited unambiguously poor fit with the data collected in this study. Conversely, the two-factor model tested for the SFNE, a model with the straightforward and reverse-scored items specified to load on two separate, correlated factors, displayed a more than sufficient degree of fit. This suggests that the SFNE does not possess a simple unidimensional structure. The value for the χ^2 statistic was significant for this two-factor model, which could be argued to suggest that the models lack sufficient fit. However, as Hair et al. (2014) point out, with large samples ($n > 250$) and a model with 12 or more indicators (items), as is the case here, a significant χ^2 value is to be expected. For this reason, the result of the χ^2 statistic should be evaluated in light of the values for the goodness-of-fit indices for the model. The value of the SRMR index for this model was substantially below Hu and Bentler's (1999) suggested cut-off (.08), indicating that it substantially reproduced the covariance relationships found in the data. The values for the TLI and CFI, while not meeting the .95 threshold, were all above .90, the point at which a model may be deemed acceptable according to early literature (Bentler, 1990). In addition, the value for the CFI was greater than .92, which Hair et al. (2014) suggests as indicating good fit in samples greater than 250 respondents. The remaining goodness-of-fit index, the RMSEA, is slightly more problematic, as the value for this index exceeded Hu and Bentler's recommendation (.06). However, the value is slightly below .80, which is the value put forth by Browne and Cudeck (1993) as indicative of acceptable fit. While this more licentious criterion comes from a somewhat dated source, there is a pattern in the general literature recently that below .06 is good and below .08 is adequate. The degree of fit exhibited by the two-factor model of the SFNE strongly suggests that scores on this instrument are not unidimensional in nature and therefore also should be interpreted with caution, a finding which corresponds to that of Nihei et al. (2018).

The results from CFAs conducted on measurement models of the SFNE would seem to suggest that a model with two underlying factors fits the scores in this dataset better than a model with a single factor. The secondary factor in this model comprised the reverse-scored items on the SFNE. This result suggests that responses to the reverse-scored items are influenced by something other than participants' degree of social evaluative anxiety, as would be the case if the instruments possessed a simple unidimensional structure. In other words, the results of the CFAs provide evidence for the existence of a degree of systematic variation caused by the reverse-scored items included on the SFNE. While some have suggested that such systematic variation can be interpreted as evidence for an additional substantial latent factor underlying the instrument, (e.g., Rodebaugh, Woods & Heimberg [2007] in regards to the SIAS), the most likely, as well as the most parsimonious, explanation for the apparent non-unitary structure displayed by the instruments in this study is the existence of a method-effect brought about by respondent confusion over the wording of the reverse-scored items (see Brown, 2003). However, regardless of the underlying cause, it is still the case that these particular items are the source of the non-random measurement error apparent in all three instruments. The fact that this error is non-random suggests that the reverse-scored items should be removed from the instruments to improve their validity, as well as their performance.

Reverse-scored items are most often included on an instrument to limit response bias, or at least make its presence easier to detect. However, as Brown (2003) points out, for this technique to be effective, the number of straightforward and reverse-scored items on an instrument should be equal in number. On the SFNE, the straightforward items outnumber the reverse-scored items by a 2:1 ratio (8 to 4 in each case), and thus the rationale for retaining the reverse-scored items as a means to limit response bias is not valid for this measure.

Secondly, as discussed in the Introduction section of this paper, previous research (e.g., Nihei et al., 2018) has shown that responses from the reverse-scored items on this instrument correlate with other measures of social anxiety to a much lower degree than either the full instrument or the straightforward items taken alone, and in some cases do not correlate at all. In addition, Nihei et al. (2018) found that a scale comprising only the straightforward items exhibited greater reliability than the full scale, and supposed that this was because of the removal of the poorly performing reverse-scored items.

Finally, and most importantly considering the context for this study, learners in the classroom, reverse-scored items tend to cause confusion in respondents with lower levels of education (Weeks et al., 2005) or who are less able readers (Marsh, 1996). While Weeks et al. (2005) found that this was most true for respondents with less than a college-level education, Marsh (1996) found similar effects among university students, as well as in 10th grade high school students. Moreover, the confusion engendered by these reverse-scored items can lead to the underestimation of the degree of social evaluative anxiety in respondents with lower levels of education (Weeks et al., 2005). Considering that working in groups is being emphasized at all levels of the Japanese education system, the possibility of underestimating the degree of social anxiety experienced by younger learners should be avoided as much as possible.

The second measure examined in this study, the SFNE-S, includes only the straightforward items from its respective parent instrument, and thus, may provide an alternative to the use of the SFNE and its troublesome reverse-scored items. As mentioned in the Introduction above, the use of this measure has been suggested in the literature concerned with the factor structure of the SFNE. For this reason, a single factor model for this measure was tested. However, the model for the SFNE-S exhibited less than satisfactory fit. While the values for three of the goodness-of-fit indices—the TLI, the CFI, the SRMR—suggested arguably adequate fit, the RMSEA value (see Table 5.4) was greater than .1, the value at which MacCallum et al. (1996) recommends for the rejection of a model. Perhaps this was because the removal of the reverse-scored items, and the corresponding confusion that they bring with them, had the result of making other sources of mis-fit in this model more apparent. On the basis of this result, it was decided to investigate the areas of local strain in the model of the SFNE-S.

As outlined above, this examination revealed the presence of a method-effect stemming from content-overlap (Brown, 2003), that is, a greater than expected covariance between items due to a similarity in wording or subject matter. More importantly, it may be the case that this is a method effect that arises more prominently in the Japanese version of the instrument than in the English version. Item 2, *I am frequently afraid of other people noting my shortcomings*, and Item 5, *I am afraid that people will find fault with me*, appear to differ slightly in their intended focus, with Item 2 more involved with an individual revealing their own inability or inadequacy in front of others through their own actions, while Item 5 seems to imply that others will criticize one's actions regardless of the actual quality of these actions. However, in the Japanese version of the FNE (Ishikawa et al., 1992), the term ketten (欠点) is used in both items. In addition, when these two items were shown to several native Japanese speaking professors and instructors with some experience in testing, all determined that these two items expressed virtually equivalent meanings. This would suggest that rather than expressing similar, though distinct, aspects of the fear of negative evaluation, as in the original FNE (Watson & Friend, 1969), these two items on the Japanese version of the SFNE (and also the FNE as well) are rewordings of an almost identical sentiment, thus engendering a method effect as outlined by Brown (2003; 2015). If these two items are in fact expressing the same sentiment, then the validity of the measure could be improved, without narrowing its operational bandwidth, by the removal of one of either of the items from the instrument.

For this reason, two alternative single-factor models—one with Item 2 removed and Item 5 retained (the first rival model), and the other with Item 5 removed and Item 2 retained (the second rival model)—were tested. Both models were found to exhibit a very high degree of fit with the underlying structure of the scores, with values for all four of the goodness-of-fit indices satisfying Hu and Bentler's (1999) recommendations, and moreover, non-significant results for the $\chi 2$ test statistic. As the two models are not nested, it was not possible to employ the $\chi 2$ difference test to determine which of the two models fit the structure of the scores more closely. While the values for the indices for the second rival model were better than those for the first, the differences were rather small, and thus it is difficult to determine if these results provide evidence for the superior fit of the second model, or if this difference is simply due to sample-specific variation present in this particular dataset.

Determining which of the two versions of the SFNE-S tested in this study provides better fit requires further testing with new datasets. However, the results of the CFAs carried out on the two rival models appear to present strong evidence that the removal of one of the two items expressing quite similar meanings (Items 2 or 5) would unquestionably improve the fit of the SFNE-S scale. This in turn suggests that a new version of the SFNE-S, lacking one of these items, be developed and the structure of this instrument tested against the dimensionality of the scores in a new dataset. This course of development would hopefully lead to an instrument which was able to generate structurally valid and reliable scores in the Japanese EFL context. In addition, this process of development, would allow researchers and practitioners to utilize the newly developed instrument with a degree of confidence due its basis in both theory and empirically derived evidence.

6. Conclusion

The instrument examined in this study, the SFNE, is extensively employed in the investigation of social evaluative anxiety. However, recent research by Nihei et al. (2018) and others has cast doubt on the unidimensional structure of the SFNE, and therefore the interpretability of its scores. This study has aimed to confirm and extend the findings of Nihei et al. (2018) by examining the factor structure of the SFNE in a sample from the Japanese university EFL learner population. In addition, this study aimed to ascertain the viability of the SFNE-S as an independent measure of social evaluative anxiety. Evidence for the validity and reliability of this scale would be an important first step in proposing the use of this scale as alternatives to the problematic SFNE.

In regards to the SFNE, this study found strong evidence that scores on this instrument lack a simple unidimensional structure, which confirmed the findings of Nihei et al. (2018). The two-factor model of the SFNE displayed good fit with the structure of the scores, providing substantial support for the hypothesis that this instrument is not a unitary measure of social evaluative anxiety, and therefore that scores generated by the SFNE in the Japanese context should be interpreted with caution.

The results from the CFAs conducted on models of the SFNE-S, on the other hand, suggest that with some adjustment, that is, the removal of one of the two items with highly similar content, this scale could serve as a useful and interpretable measure of social evaluative anxiety. It should be noted here that establishing the structural validity of an instrument is a cumulative process and thus additional research is needed, not only to determine which of the two problematic items should be removed from the instrument, but also to provide further evidence for or against the structure proposed in this study, and in an a priori rather than a posteriori test.

More generally, and in a similar manner to the findings of Brown (2003), this study's findings highlight the importance of investigating the existence of a range of possible method effects when evaluating the results from CFAs carried out on structural models. As Brown points out, non-random measurement error (i.e. the influence of method effects) is not unusual, however, many studies report only the values of those indices employed to estimate global fit, neglecting to examine other possible issues in measurement models, as is often recommended (e.g., Byrne, 2016; Kline 2011).

The fact that this study has examined scores from only one sample from the target population, Japanese university EFL learners, may limit the generalizability of its results. However, this is a limitation which is applicable to much other similar research in the literature, and in social science research more generally. Repeated sampling of the target population would not only serve as a means of overcoming this limitation, but would also contribute valuable evidence to the cumulative process of determining the validity of the SFNE and its variants among Japanese university EFL learners.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41*(12), 1411-1426.

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.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.

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

Collins, K. A., Westra, H. A., Dozois, D. J., & Stewart, S. H. (2005). The validity of the brief version of the Fear of Negative Evaluation Scale. *Journal of Anxiety Disorders, 19*(3), 345-359.

Curtiss, J., Klemanski, D. H., Andrews, L., Ito, M., & Hofmann, S. G. (2017). The conditional process model of mindfulness and emotion regulation: An empirical test. *Journal of Affective Disorders, 212*, 93-100.

Essau, C. A., Sasagawa, S., Chen, J., & Sakano, Y. (2012). Taijin kyofusho and social phobia symptoms in young adults in England and in Japan. *Journal of Cross-Cultural Psychology, 43*(2), 219-232.

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.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis: Pearson new international edition*. Essex, UK: Pearson Education Limited.

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.

石川利江 [Ishikawa Rie], 佐々木和義 [Sasaki Kazuyoshi], & 福井至 [Fukui Itaru]. (1992). 社会的不安尺度 FNE・SADS の日本版標準化の試み [Standardization of Japanese version of FNE and SADS]. 行動療法研究 [*Japanese Journal of Behavioral Therapy*], *18*(1), 10-17.

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.

Jöreskog, K. G., & Sörbom, D. (1993) *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

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, UK: Multilingual Matters.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Leary, M. R. (1991). Social anxiety, shyness, and related constructs. In J.P. Robinson, P.R. Shaver, & L. W. Wrightsman (Eds.), *Measures of personality and social psychological attitudes: Measures of social psychological attitudes,* (Vol. 1, pp. 161-195). San Diego, CA: Academic

Press

Leeming, P. (2011). Japanese high school students' use of L1 during pair-work. *International Journal of Applied Linguistics*, *21*(3), 360-382.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149.

前田香 [Maeda Kaori], 関口真有 [Sekiguchi Mayu], 堀内聡 [Horiuchi Satoshi], Weeks, J. W. & 坂野雄二 [Sakano Yuji]. (2015). Fear of Positive Evaluation Scale 日本語版の信頼性と妥当性の検討 [Reliability and Validity of a Japanese Version of the Fear of Positive Evaluation Scale]. 不安症研究 [*Anxiety Disorder Research*], *6*(2), 113-120.

Maeda, S., Shimada, H., Sato, T., Tashiro, K., & Tanaka, Y. (2017). Translation and initial validation of the Japanese version of the self-beliefs related to social anxiety scale. *Psychological Reports, 120*(2), 305-318.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810-819.

Moriya, J., & Tanno, Y. (2011). The time course of attentional disengagement from angry faces in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 42*(1), 122-128.

二瓶正登 [Nihei Masato], 荒井穂菜美 [Arai Honami], 前田香 [Maeda Kaori], 青木俊太郎 [Aoki Shuntaro], 土屋垣内晶 [Tsuchiyagaito Aki], 岩野卓[Iwano Suguru], …坂野 雄二 [Sakano Yuji]. (2018). Fear of Negative Evaluation Scale 日本語短縮版の因子構造, 信頼性および妥当性の再検討 [Reassessment of factor structure, reliability, and validity of the Short Fear of Negative Evaluation scale for the Japanese]. 不安症研究 [*Anxiety Disorder Research*], *10*(1), 54-63.

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

Okajima, I., Kanai, Y., Chen, J., & Sakano, Y. (2009). Effects of safety behaviour on the maintenance of anxiety and negative belief social anxiety disorder. *International Journal of Social Psychiatry, 55*(1), 71-81.

Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: The factor structure and item properties of the original and Brief Fear of Negative Evaluation scale. *Psychological Assessment*, *16*(2), 169-181.

Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). The reverse of social anxiety is not always the opposite: The reverse-scored items of the Social Interaction Anxiety Scale do not belong. *Behavior Therapy*, *38*(2), 192-206.

笹川智子 [Sasagawa Satoko], 金井嘉宏 [Kanai Yoshiro], 村中泰子 [Muranaka Yasuko], 鈴木伸一 [Suzuki Shinichi], 嶋田洋徳 [Shimada Hironori] & 坂野雄二 [Sakano Yuji]. (2004). 他者からの否定的評価に対する社会的不安測定尺度 (FNE) 短縮版作成の試み: 項目反応理論による検討 [Development of a short Fear of Negative Evaluation scale for Japanese using item response theory]. 行動療法研究 [*Japanese Journal of Behavior Therapy*], *30*(2), 87-98.

Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin, 92*(3), 641-669.

城月健太郎 [Shirotsuki Kentaro], 笹川智子 [Sasagawa Satoko], & 野村忍 [Nomura Shinobu]. (2007). ネガティブな反すうが社会不安傾向に与える影響 [The effects of negative rumination on social anxiety]. 健康心理学研究 *[Japanese Journal of Health Psychology]*, *20*(1), 42-48.

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

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., & Liebowitz, M. R. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological assessment*, *17*(2), 179-190.

Xethakis, L. (2020) *The Psychometrics of Affective, Cognitive and Social Factors Impacting Group-Based Learning Activities* [Unpublished doctoral dissertation]. Kumamoto University

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: The Short Fear of Negative Evaluation Scale (SFNE; Sasagawa et al., 2004)

	Item
1.	人がなんと思おうと，どうということはないとわかっていても，自分のことを人がどう思か気になる。I worry about what people will think of me even when I know it doesn't make any difference.

2. 他の人が私の欠点に気づくのではないかとしばしば心配する。I am frequently afraid of other people noticing my shortcomings.
3. 他の人が自分のことを認めてくれなくても， あまり気にならない。The disapproval of others would have little effect on me. *
4. どんな印象を人に与えているか，ほとんど気にしない。I rarely worry about what kind of impression I am making on someone. *
5. 人に自分の欠点を，みっけられるのではないかと心配だ。I am afraid that others will find fault with me.
6. 誰かと話しているとき，その人が自分のことをどう思っているか心配だ。When I am talking to someone, I worry about what they may be thinking of me.
7. 自分がどんな印象を与えているのかいつも気になる。I am usually worried about what kind of impression I make.
8. 他の人が私のことを価値がないと思うのではないかと心配だ。I worry that others will think I am not worthwhile.
9. 他の人が私のことをどう思うかはほとんど気にならない。I worry very little about what others may think of me.*
10. 他の人が私のことをどう思っているか，気にしすぎると思うことがときどきある。Sometimes I am too concerned with what other people may think of me.
11. 他の人が私をどう思っているか気にかけない方である。I am often indifferent to the opinions others have of me. *
12. 私の友達が自分をどう思っているかをあれこれ考えてしまう。I brood about the opinions my friends have about me.