Examining Motivation to Study Abroad Among Japanese Undergraduates

Aaron C. Sponseller

日本の大学生の海外留学の動機を探る

スポンセラー　アーロン

Abstract

Attempts to develop an instrument (s) for measuring study abroad motivation have often lack robust evidence of reliability, and have not reliably shown the capacity to distinguish the motivational profiles of students who choose to engage in study abroad and those who do not. This study is an initial attempt to address these issues. A Japanese version of Anderson and Lawton's (2015) Motivation to Study Abroad (MSA) scale was operationalized with Japanese undergraduates (N = 251). Rasch rating scale modeling indicated the MSA scales are reliable but generally too easily endorsed. While significant differences were observed between schools and gender for each of the four MSA scales, no significant difference between sojourners and non-sojourners was detected. Finally, MSA scale scores were not capable of predicting which students would study abroad and which would not. Implications and future research directions are offered.

抄　　　録

　海外留学のモチベーションを測定するツール開発の試みは、しばしば信頼性の確固たる証拠を欠き、海外留学を選択する学生とそうでない学生のモチベーションプロファイルを区別する能力を確実に示すことができなかった。本研究は、これらの問題を解決するための最初の試みとして、Anderson and Lawton（2015）の Motivation to Study Abroad（MSA）スケールの日本語版を、日本の大学生（N = 251）を対象に運用した。ラッシュ評定尺度モデリングにより、MSA 尺度は信頼性が高いが、一般的には簡単に支持されすぎていることが示された。MSA の 4 つの尺度それぞれについて、学校や性別による有意差が見られたが、留学生と非留学生の間には差が見られなかった。最後に、MSA 尺度の得点は、どの学生が留学するか、しないかを予測することはできなかった。以上の結果を踏まえて、今後の研究の方向性について提案する。

Background

Study Abroad Trends

Prior to the emergence of COVID-19 global pandemic, study abroad trends among university undergraduates globally showed a rapid growth. Japan was no exception to this trend. Study abroad participation has also, for many decades, been markedly higher among female undergraduates than their male counterparts.

According to data from the Japan Student Services Organization (JASSO, n.d.), total study abroad participation increased from under 25,000 in 2004 to well over 100,000 in 2019. Figure 1 shows the year-to-year growth in total study abroad participation among Japanese undergraduates according to historical data available from JASSO.
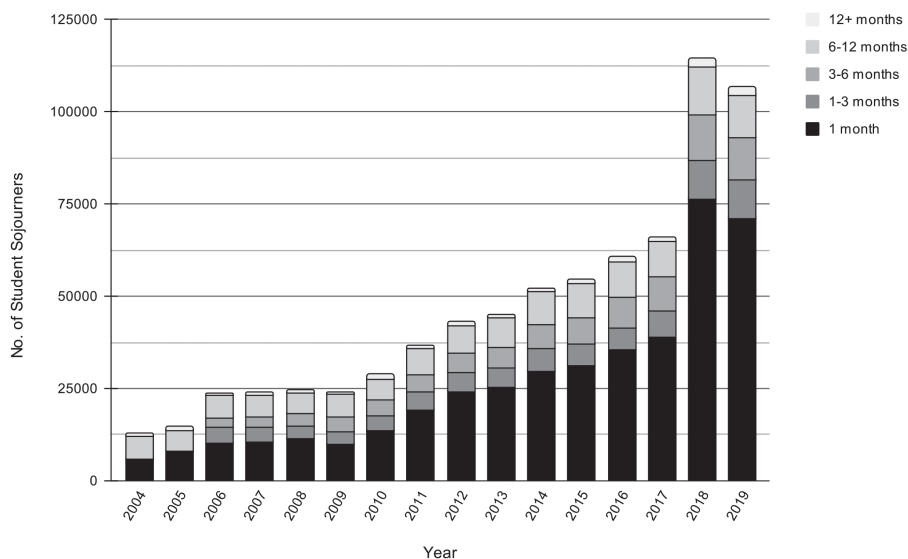
Figure 1   JASSO total study abroad participation rates by year and program duration

Another long-established trend globally, and to which Japan is no exception, is that female university students participate in study abroad programs at a significantly higher rate than their male counterparts (Dessoff, 2006). Work by Kinginger (2009) asserts that study abroad has become feminized. Recent work by Williams (2020) challenges the

narrative of study abroad as a feminine act, focusing on Japanese male sojourners' ($N =$ 25) retrospective accounts of their time abroad and the critical role English language development played in their time abroad and overall identity formation post-sojourn. There is, however, zero doubt that Japanese women have been engaging in study abroad at a greater rate for at least the past two decades. Figure 2 displays publicly available JASSO (n.d.) data from 2004 through 2019. While study abroad participation among men has increased in total by a substantial margin over that time, women's participation has always exceeded it by a significant margin.
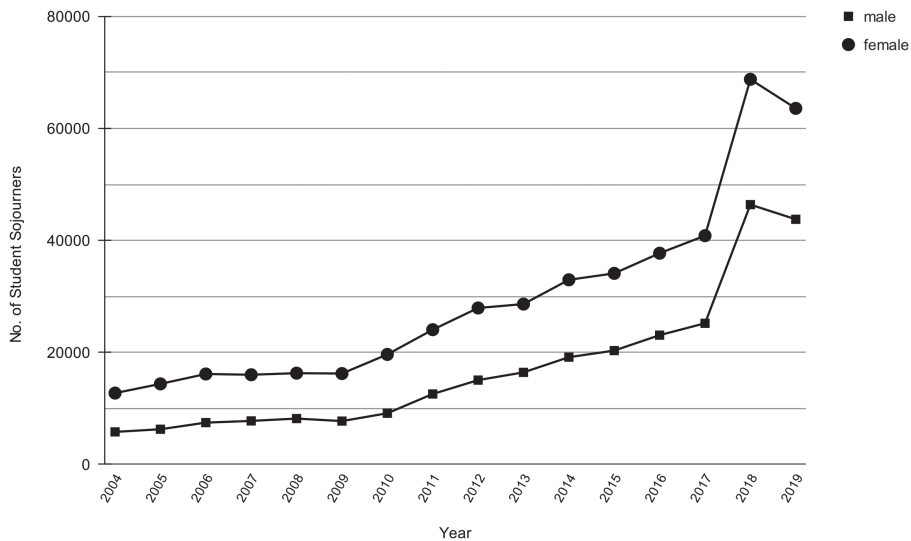
Figure 2　JASSO total study abroad participation rates by year and gender

Furthermore, JASSO data (n.d.) suggests that participation in study abroad varies quite widely by academic major, with STEM majors participating at a far lower rate than those students majoring in the soft or social sciences. This aligns with claims from NAFSA (2008) that curricular restrictions in STEM majors inhibit study abroad participation. Salisbury et al. (2009) found no between-major differences in student interest to study abroad. Both Li et al. (2013) and Salisbury et al. (2010) suggest study abroad marketing should differentiate their approach to targeting men and women to participate in study abroad, with efforts to recruit male participants being wise to emphasize the possibility of achieving something or increasing eventual employability. Salisbury et al. (2010) also suggest reducing the early pressure on male students to select a major could facilitate an increase in their study abroad participation rates.

Motivation to Study Abroad

The literature on study abroad motivation paints a rather confusing, multifaceted, and unresolved picture. The allure of a multilingual, parsimonious, and multifaceted instrument capable of measuring aspects of would-be sojourner motivation is compelling for at least two reasons. First, a multifaceted instrument would be in-line with Self-Determination Theory (SDT; Deci & Ryan, 2012). SDT recognizes motivation is not as a unitary phenomenon but dynamic at the individual, group, and contextual levels-and these motivational forces directly inform subsequent behaviors. Second, any such instrument would ideally also hold some capacity to predict study abroad related behavior. Of most immediate utility would surely be whether the motivational scale(s) could reliably predict whether or not a would-be sojourner becomes an *actual* sojourner. Additionally, any cross-culturally validated instrument measuring aspects of study abroad motivation would potentially enable study abroad researchers and program administrators to explore a myriad of subsequent research questions. For example, do certain motivational profiles predict what types of programs students will choose, how long they will stay abroad, or the degree to which they engage with their local/host community whilst abroad? Additionally, what is the relationship between aspects of study abroad motivation and various learning outcomes or sojourner growth in areas such as intercultural sensitivity or cross-cultural understanding? In short, any instrument that reliably measures the various (de)motivational currents within would-be student sojourners opens up vast fresh territory for empirical exploration in the field of study abroad.

Several studies have attempted to develop psychometrically validated scales measuring aspects of sojourner motivation. Three relatively recent such studies (Anderson & Lawton, 2015; Nyaupane et al., 2010; Sanchez et al., 2006) have attempted to develop instruments which measure sojourner motivation. Anderson and Lawton's (2015) study built its item inventory off the work of Nyaupane et al. (2010) and Sanchez et al. (2006) and makes a straightforward argument that the underlying factor structure for each of these respective instruments is quite similar. While the MSAs four scales of World Enlightenment (WE), Personal Growth (PG), Career Development (CD), and Entertainment (EN) are indeed represented in each of the three aforementioned studies, additional motivational factors have been suggested by other researchers. The work of Kitsantas (2004) and Wiers-Jenseen (2003), for example, suggests the four common motivational factors for sojourners are cross-cultural experience, academics, future careers, and family heritage. Much more recently, Aresi et al. (2017) developed a Multidimensional Motivation to Study Abroad Scale (MMSAS) composed of nine scales: personal growth, better academic knowledge, others' expectations, learning or improving foreign language skills, cross-cultural interest, get away from home environment, career perspectives, search

or independency, and leisure. Validity evidence was presented for the English, Dutch, French, German, Italian, and Spanish versions of the MMSAS.

Somewhat problematically, the theory underpinning the various study abroad motivation instruments described above is extremely varied. As an example, the MSA developed by Anderson and Lawton (2015) builds upon and attempts to consolidate the work of Sanchez et al. (2006) and Nyaupane et al. (2011), yet these two studies theoretical approaches which appear to lack any coherent overlap. Sanchez et al. (2006) approached motivation to study abroad from the theoretical framework of Expectancy Theory (Vroom, 1964; Wang, 2004). Expectancy Theory, as applied to motivation and study abroad, contends that would-be sojourners will make an effort to achieve performance and reward they perceive as valuable. In Sanchez et al.'s (2006) model, they also drew a distinction between motivation to study abroad and intent to study abroad. In their model, intent to study abroad was rooted in motivations to study abroad that are moderated by familial, financial, psychological, and social barriers. Approaching the issue via Gnoth's (1997) theoretical model of tourism motivation and Hull's (1943) Drive Theory, Nyaupane et al. (2011) explored the influence of sojourner motivations, past travel experience, and social ties in selecting destinations for study abroad programs. They found that academic motivation and social ties in the destination country were the most critical factors in sojourner destination selection. The theoretical underpinning of motivational studies varies widely.

Concerning instrument validity and reliability, evidence supporting the MSA in both English and Japanese is limited. Anderson and Lawton reported confirmatory factor analysis fit indices showing the MSA had adequate model fit (CFI = 0.91; TLI = 0.90; RMSEA = 0.074; SRMR = 0.062), and Cronbach's alpha scores for WE (0.91), PG (0.86), CD (0.90), and EN (0.81) indicated generally good scale reliability. In the context of Japan, recent preliminary work by Sponseller (2020) with Japanese undergraduates (N = 165) reported lower but still acceptable alphas for WE (0.87), PG (0.86), CD (0.80), and EN (0.80). After fitting a model which allowed several items' errors to covary, Sponseller (2020) reported CFA fit of $\chi^2(221) = 100.735$, $p < .001$; χ^2/df = 1.937, CFI = 0.872; SRMR = 0.088; RMSEA = 0.081, indicating poorer fit in the Japanese context than in the original study.

In terms of practical utility, the instruments developed by Anderson and Lawton (2015), Nyaupane et al. (2010), and Sanchez et al. (2006) have thus far not been shown to connect motivation scale scores to any type study abroad outcome. The only study attempting to tie MSA scores to study abroad outcomes (Anderson et al., 2015) failed to connect MSA scales scores with any pre-post change in Global Perspective Inventory (GPI; Braskamp et al., 2010) scores among US undergraduate sojourners ($N = 355$). As such, the questions Anderson and Lawton (2015) closed their study with appear to

remain unanswered. Among those questions are whether differing MSA profiles lead to differences in program destinations, durations, housing arrangements, how sojourners spend their out-of-class time whilst abroad, and learning outcomes.

Research Questions

Given the relatively underwhelming evidence of reliability currently available for the MSA, as well as the absence of evidence the MSA can establish differences in sojourn participation by gender or predict which students will ultimately decided to engage in a sojourn abroad, in this study I explore the following research questions:

1. To what extent does the MSA appear to be a valid and reliable instrument in the Japanese context?
2. In what ways do aspects of study abroad motivation, as measured by the MSA, differ between schools, gender, and those who study abroad versus those who do not study abroad?

Methods

Participants and Context

Participants were drawn from two universities in Western Japan: Momiji University and Sakura University (both pseudonyms). Momiji University is a large national university consistently ranked as one of the top 10-15 universities in Japan by *hensachi*. Sakura University is a small, private women's university and junior college that is on the other end of the continuum based on *hensachi*. *Hensachi* is a score that expresses how far from the statistical mean a student who has been admitted to the university scores on the entrance exam. At both universities all participants were recruited via convenience sampling. At each university the appropriate ethical protocols were followed and I had permission to gather data from students. Participants were informed that their participation was entirely voluntary, did not influence their course grade, and that their responses would remain anonymous. Participants were therefore offered a gift card of 1000 yen to complete a survey battery of which the MSA was one part. All participants provided written informed consent. Upon obtaining consent, participants were emailed a link to complete the survey battery.

All MSA data analyzed in this study was collected between 2017 and 2019, prior to the outbreak of the COVID-19 global pandemic. This ensures COVID-19 itself, government reactions to handling the pandemic, and all manner of COVID-19-related issues were not influencing students' responses to the MSA. Data was collected from three separate

cohorts of students. The first cohort was at Momiji University, and data was collected between the end of the spring semester in 2017 and the beginning of the fall semester 2017. The second cohort was also drawn from Momiji University, with data collected between the end of the 2017 fall semester (e.g. January 2018) and the beginning of the 2018 spring semester. The third cohort was drawn from Sakura University, with data collected between the end of the spring semester in 2019 and the beginning of the fall semester 2019. Each of these cohorts consisted of both students who were going to participate in short-term study abroad programs between the semesters, and control group of students who were not participating in study abroad during the semester break.

A total of 263 students provided their written informed consent to participate and completed all instruments. After initial data screening and analysis, including Rasch rating scale validation procedures described subsequently, a final dataset of 251 participants (78.5% female) was used for all analyses in this study. Of the 251 participants, 156 were from Momiji University and 95 were from Sakura University. At Momiji University 102 of the participants (65.4%) identified as female, and 91 (58.3%) were not planning to study abroad during the course of data collection. At Sakura University, 100% of the population identified as female and of the 95 total participants 35 (36.8%) were not planning to study abroad during the course of data collection.

Instrumentation

Motivation to Study Abroad (MSA) Scale. The MSA consists of 23 items and four subscales: World Enlightenment (7 items), Personal Growth (6 items), Career Development (5 items), and Entertainment (5 items). An example of an item for World Enlightenment would be *become acquainted with people different from me*. The Personal Growth scale included items like *grow as a person*. Career Development included items such as *build my resume*. An example of an item measuring Entertainment as a motivation for studying abroad is *experience the local nightlife (clubs, bars, etc.)*. See Appendix A for the full set of English and Japanese MSA items, divided by subscale.

The response scale for the MSA used in this study differed from the original work of Anderson and Lawton (2015). While the original MSA used a five-point response scale ranging from *not at all important to absolutely essential*, previous work on the MSA by Sponseller (2020) indicated that Japanese responses to the MSA tended to skew positive. Additionally, there is response scale research suggesting that the middle response category in odd-numbered Likert response scales tends to not serve as a neutral category but as an "it depends" category (Kulas & Stachowski, 2009, 2013). While the middle response category on the original MSA was important and not a potentially more problematic descriptor such as neutral, my experience with response scales in the

Japanese context led me to develop a six-point response scale as follows:

　　1 – not at all important; 全く重要でない

　　2 – not important; 重要でない

　　3 – not very important; あまり重要でない

　　4 – a little important; 少し重要である

　　5 – important; 重要である

　　6 – very important; 非常に重要である

　　The original English MSA items were translated into Japanese by three individuals before being evaluated by a fourth individual. All translators were at Momiji University and translated the items from English into their L1 (Japanese). Two were doctoral students studying English language pedagogy, and the third held a PhD from a major research university in the United States. These translations were then shared with a professor of English language education at Momiji University who determined which translations would most appropriately convey the meaning of the original English items in the Japanese context.

Analysis

　　For research question one, the same procedures as Anderson and Lawton (2015) are followed. Descriptives, alphas, and boxplots illustrating MSA profiles based on school, gender, and participation in study abroad or not. Next, MSA scores are transformed into interval level data using Rasch rating scale modeling. Evidence concerning rating scale utility, as well as item and person reliability and unidimensionality, are then presented.

　　There are four reasons why Rasch analysis is appropriate in this case. First and foremost, Rasch transforms ordinal data into interval scale data, whereas the use of untransformed (e.g. raw) Likert-style data means treating ordinal data as if it were interval data (Andrich, 1978; Masters, 1982; Wolfe & Smith Jr, 2007). Interval data assists in making accurate comparisons between respondents, items, and their respective abilities and difficulties. This is precisely what MSA scale scores are intended for in research questions two and three. Second, Rasch modeling assists when confirming a hierarchy of item endorsability. Items range in terms of difficulty or endorsability and, therefore they contribute differentially to the overall measurement of a construct. Rasch analysis places all respondents and items around a single logit-scale line, and the separation between persons and items is thus presented in precise order of ability/difficulty. Third, Rasch analysis provides fit statistics not available in classical test theory approaches. These fit statistics include infit, outfit, mean square (MNSQ), and standardized fit statistics (ZSTD). Fourth, Rasch analysis can assess the dimensionality of the construct being measured. Bond and Fox (2015) have identified two primary problems with relying solely on factor

analytic approaches when offering evidence of construct validity. First, such approaches assume that the data are interval in nature, when they are actually ordinal. Second, factor analytic approaches do not provide item level fit statistics, a shortcoming which limits the degree to which we can determine how well items fit/misfit the construct under investigation. While internal reliability measures such as Cronbach's alpha and MacDonald's Omega are frequently used to provide evidence of unidimensionality, these measures really offer evidence of item correlation (Embretson & Reise, 2000; Waugh & Chapman, 2005).

For research question two, a series of Bonferroni-adjusted Welch's t-tests and Mann-Whitney U-tests are conducted to look for significant differences in MSA scale scores based on school, gender, and participation in study abroad or not.

Results

Research Question 1. See Table 1 for descriptive statistics and alphas.

**Table 1 Descriptives and reliability of raw MSA scale responses**

Scale	# of items	N	M	SD	Cronbach's alpha	Anderson & Lawton's (2015) alpha
World Enlightenment	7	251	5.45	0.11	0.89	0.91
Personal Growth	6	251	5.27	0.21	0.84	0.86
Career Development	4	251	4.57	0.47	0.79	0.89
Entertainment	5	251	3.10	0.33	0.79	0.82

Of particular note in Table 1 is that the mean scores for WE and PG scales are quite high, suggesting student responses skewed positive. Such findings were similarly found in Anderson and Lawton's (2015) original study, as well as Sponseller's (2020) study with a smaller sample of students.

MSA scale scores were then explore visually using boxplots grouped by school and study abroad status. See Figure 3.

As Figure 3 clearly illustrates, the MSA scales appear to follow a consistent trend at both Momiji and Sakura University. At each university, the mean scores for World Enlightenment and Personal Growth scales are both extremely high. Scale scores for Career Development are only slightly less high, while scores for Entertainment are, on average, comparatively low. Study abroad status did not appear to distinguish between participants either within or between schools on any of the scales.

In an effort to increase the separation between participants on the MSA scales-particularly the scales of World Enlightenment and Personal Growth which were skewed so positively-MSA scores were then transformed into interval level data via Rasch rating
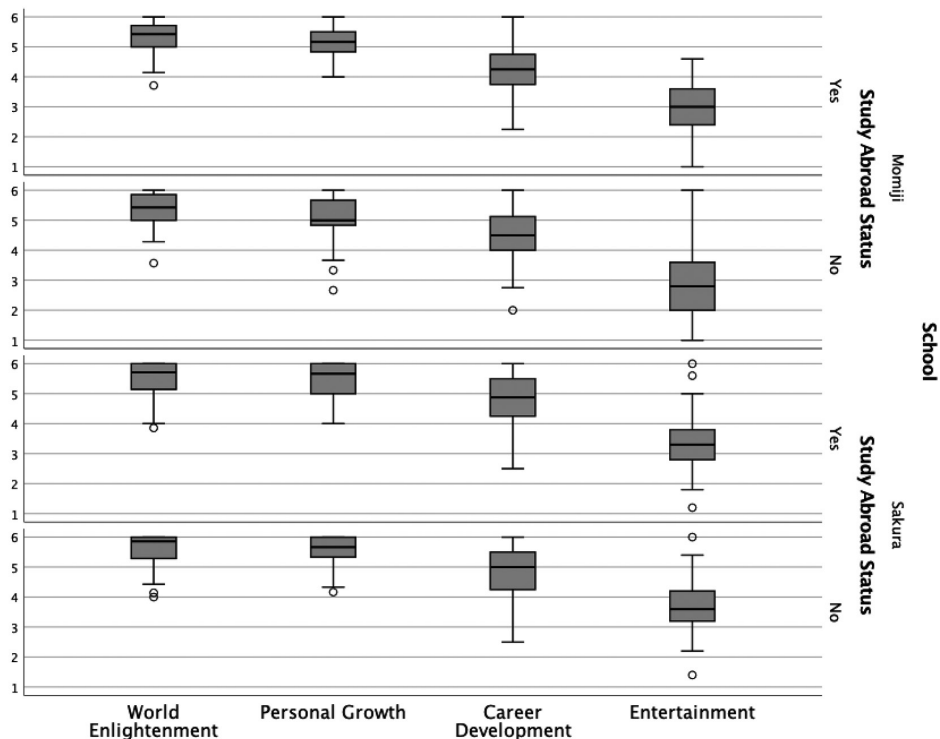
Figure 3 Boxplot of MSA scale scores by school and study abroad status

scale modeling (Rasch, 1960) using version 4.3.0 of the software WINSTEPS (Linacre, 2018). Raw data from the MSA was entered into separate command files. For each scale, evidence concerning extreme outliers, response scale functioning, item fit statistics, and dimensionality will now be provided.

Extreme outliers are those participants whose response to items were considered so systematically unpredictable as to constitute either a participant who completed the instruments in bad faith or simply did not understand the items. Upon initial inspection of the data, six participants provided the same response for all items (e.g. providing all 1s or all 6s for each of the 23 items on the instrument). These participants were removed from subsequent analyses. Next, I checked the person Infit and Outfit MNSQ statistics, looking for those individuals whose were among the top 5% of misfitting people on multiple scales. On my first pass through the data I found one individual who was among the three most misfitting persons on all scales. I removed this individual from subsequent analyses and ran the data again. I then found five additional individuals who were among the top 5% of most misfitting persons on at least two scales. Therefore, I removed these individuals from subsequent analyses. In total, 12 persons considered extreme outliers were removed from the dataset, leaving a final participant pool of 251 individuals. The scales of World

Enlightenment and Personal Growth were somewhat problematic when attempting to identify outliers, as roughly 20% of the participants scored a maximal measure on each of these scales. As I do not believe 20% of the participants were providing dishonest responses, and because 20% of the population cannot, almost by definition, be considered outliers, I kept these individuals for all subsequent analyses.

The following response scale functioning criteria were used as a general guideline. Criteria one through five come from Linacre (1999), while criterion six comes from Elwood (2011).

1. A minimum of 10 observations per response scale category
2. A normal response distribution, peaking in the middle
3. Average category measures increase sequentially
4. Outfit MNSQ statistics range from .8 to 1.4
5. Rasch-Andrich thresholds increase sequentially
6. Rasch-Andrich thresholds are .59 logits apart

As Table 2 illustrates, response scale functioning for the MSA scales was problematic. For three of the scales (World Enlightenment, Personal Growth, and Career Development), response categories 1 (not at all important) and 2 (not important) were so sparsely used that they were collapsed with response category 3 (not very important) in order to meet the minimum category use threshold of 10 responses per category. World Enlightenment and Personal Growth also exhibited a clear skew towards the positive response categories, with 83% and 89% of responses falling into response categories 5 (important) and 6 (very important) on those scales, respectively. Thus the data appear to violate the criteria 2 from above, as responses did not peak in the middle categories.

The response scale functioning for the MSA scale of Entertainment suffered from the opposite end, with category 6 (very important) used so little it was collapsed with category 5 (important) in order to meet minimum criteria.

The range of item infit and outfit MNSQ statistics were then checked. The criteria set by Fisher (2007), which are as follows, were used as general guidelines:

1. From < .33 to > 3.0 indicates poor fit
2. From .34 to 2.9 indicates fair fit
3. From .5 to 2.0 indicates good fit
4. From .71 to 1.4 indicates very good fit
5. From .77 to 1.3 indicates excellent fit

One item on the Career Development scale, MSA17 (*gain in-depth knowledge in my chosen field*) exhibited what I felt was unacceptable Infit MNSQ (1.89) and outfit (1.94), and was subsequently cut from further analysis. This item was so infrequently endorsed that it appeared no participants viewed it as a motivating factor. This makes sense considering

Table 2   Rasch category statistics of the six-point rating scale for the MSA scales

Category and descriptor	Count (%)	Infit MNSQ	Outfit MNSQ	Andrich threshold	Category measure
World Enlightenment					
3 not very important[1]	46 (3%)	0.97	0.96	NONE	(-3.33)
4 a little important	217 (14%)	1.00	1.03	-2.14	-1.20
5 important	520 (35%)	0.94	0.95	-0.20	1.11
6 very important	723 (48%)	1.03	1.02	2.33	(3.49)
Personal Growth					
3 not very important[2]	18 (1%)	1.29	1.33	NONE	(-4.39)
4 a little important	166 (9%)	1.04	1.04	-3.25	-1.73
5 important	636 (36%)	0.93	0.92	-0.20	1.63
6 very important	937 (53%)	0.98	0.98	3.45	(4.57)
Career Development					
3 not very important[3]	169 (17%)	1.11	1.10	NONE	(-3.60)
4 a little important	295 (29%)	0.99	1.05	-2.44	-1.22
5 important	296 (29%)	0.97	0.92	0.04	1.24
6 very important	244 (24%)	0.89	0.90	2.40	(3.56)
Entertainment					
1 not at all important	157 (13%)	0.97	0.97	NONE	(-3.11)
2 not important	235 (19%)	0.89	0.89	-1.78	-1.44
3 not very important	418 (33%)	0.94	0.95	-0.90	0.00
4 a little important	272 (22%)	1.01	0.99	0.90	1.43
5 important[4]	173 (14%)	1.14	1.13	1.79	(3.12)

Notes:

[1,2,3] due to low use of categories 1 (*not at all important*) and 2 (*not important*), they were collapsed with category 3 (*not very important*).

[4] due to low use of category 6 (*very important*), it was collapsed with category 5 (*important*).

the study abroad programs available to participants at these universities are not geared toward the acquisition of specialist knowledge but toward general cultural exposure or language learning. Of the 22 remaining items, 21 met Fisher's (2007) criteria of either very good or excellent fit. One item on the Personal Growth scale (MSA4: *become acquainted with people different from me*) failed to meet the criteria of good fit.

Dimensionality for all measures was checked by using a Rasch principal components analysis (PCA) of item residuals. As explicit criteria for checking Rasch PCA results for polytomous instruments vary in the research literature, I investigated dimensionality by checking the first contrast of all instruments to explore the possible presence of an additional dimension present among the item residuals. None of the first contrasts held eigenvalues over 2.0, a common threshold for detecting the potential presence of an additional dimension, therefore unidimensionality was confirmed for the MSA scales.

Item and person reliability are reported in Table 3. Item reliabilities were all robust. Item separation was likewise adequate, and for CD in particular the separation was quite good. Person reliabilities, which are always lower than item reliabilities, were acceptable. Person separation ranged from 1.57 (WE) to 1.94 (CD). Overall, the MSA scales appear reliable but are underwhelming in terms of separating people.

Table 3　Item and person reliability and separation

Scale	Item reliability	Item separation	Person reliability	Person separation
World Enlightenment	0.97	5.52	0.71	1.57
Personal Growth	0.95	4.30	0.76	1.77
Career Development	0.99	9.91	0.79	1.94
Entertainment	0.96	5.22	0.77	1.81

Finally, correlations and confidence intervals were calculated using the Rasch-transformed MSA scale scores. Figure 4 reports the correlations and displays
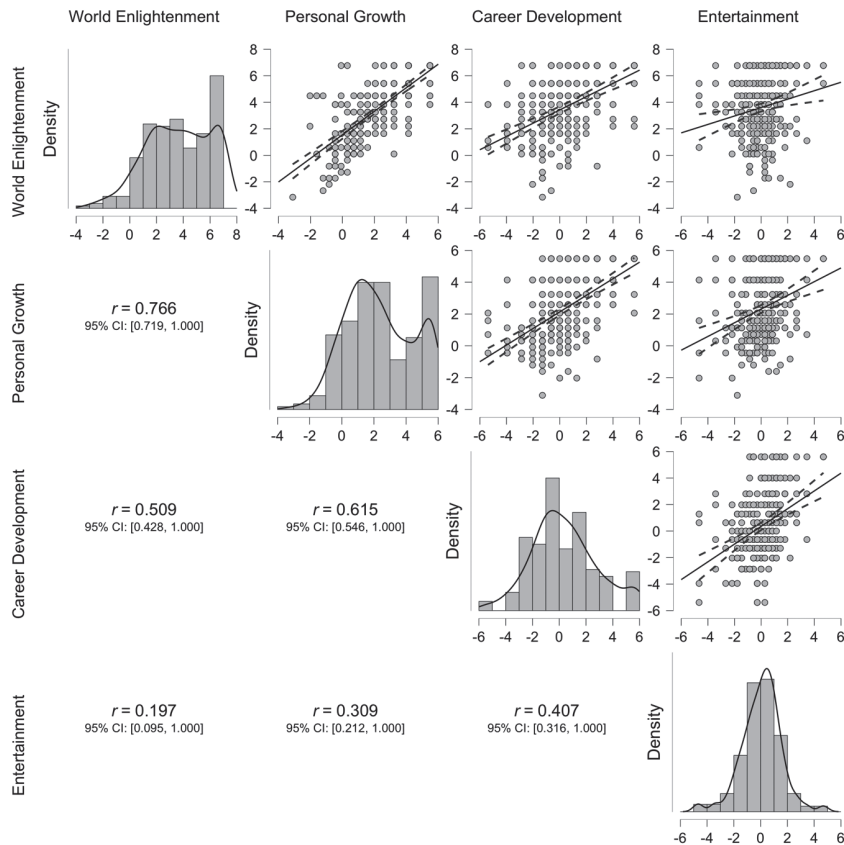
Figure 4　Correlation scatterplot for Rasch-transformed MSA scale scores
Note: all correlations significant at $p < .001$

scatterplots. Of note is that all correlations between MSA scales were significant at p < .001. Correlation between WE and PG, the two scales which exhibited strong ceiling effects, was particularly high. The scale of EN, while significantly correlated with WE, PG, and CD, had the smallest correlation coefficients.

Research Question 2. The Rasch-transformed data were input into the statistical software JASP (JASP Team, 2021) in order to conduct t-tests for group differences on MSA scale scores based school, gender, and study abroad status. Initial exploration of the data by means of a Shapiro-Wilk test indicated the MSA data was not normally distributed. See Table 4 for all Shapiro-Wilk test results.

Table 4   Tests of normality (Shapiro-Wilk) and equality of variances (Levene) for MSA scales by school, gender, and sojourner status

			Shapiro-Wilk	
			statistic	sig.
School	WE	Momiji	0.958	< .001
		Sakura	0.871	< .001
	PG	Momiji	0.960	< .001
		Sakura	0.886	< .001
	CD	Momiji	0.981	0.027
		Sakura	0.948	< .001
	EN	Momiji	0.972	0.003
		Sakura	0.958	0.004
Gender	WE	Female	0.921	< .001
		Male	0.964	0.101
	PG	Female	0.932	< .001
		Male	0.951	0.027
	CD	Female	0.971	< .001
		Male	0.930	0.004
	EN	Female	0.968	< .001
		Male	0.965	0.111
Study Abroad Status	WE	No S.A.	0.935	< .001
		S.A	0.940	< .001
	PG	No S.A.	0.941	< .001
		S.A	0.931	< .001
	CD	No S.A.	0.972	0.009
		S.A	0.963	0.002
	EN	No S.A.	0.980	0.054
		S.A	0.953	< .001

Though the sample size ($N = 251$) was large enough to arguably ignore this violation both the non-parametric Mann-Whitney U-test and Welch's t-test results will be reported

Table 5   Welch's *t*-test and Mann-Whitney *U*-test results by school, gender, and study abroad status

		Test	Statistic	df	p
School	WE	Welch	-2.863	174.572	0.005
		Mann-Whitney	5564.0		< .001
	PG	Welch	-5.021	185.609	< .001
		Mann-Whitney	4783.5		< .001
	CD	Welch	-4.462	177.034	< .001
		Mann-Whitney	5092.5		< .001
	EN	Welch	-5.106	221.077	< .001
		Mann-Whitney	4740.5		< .001
Gender	WE	Welch	3.594	103.174	< .001
		Mann-Whitney	6910.0		< .001
	PG	Welch	6.241	108.014	< .001
		Mann-Whitney	7753.0		< .001
	CD	Welch	5.518	117.149	< .001
		Mann-Whitney	7476.0		< .001
	EN	Welch	1.754	80.008	0.083
		Mann-Whitney	6202.5		0.061
Study Abroad Status	WE	Welch	0.136	248.724	0.892
		Mann-Whitney	7893.5		0.975
	PG	Welch	-0.149	247.838	0.882
		Mann-Whitney	7830.5		0.939
	CD	Welch	0.859	247.006	0.391
		Mann-Whitney	8435.5		0.328
	EN	Welch	-0.594	245.372	0.553
		Mann-Whitney	7594.0		0.625

here. See Table 5 for all results displayed by school, gender, and study abroad status.

Differences between Momiji and Sakura University were examined first. Both Welch's *t*-test and Mann-Whitney U-test results were statistically significant for all MSA scales. Results suggest mean scores on all MSA scales are systematically different between students at Momiji University and Sakura University. Inspection of distribution plots indicated Sakura university students' mean scores were higher across all MSA scales compared to their counterparts at Momiji University.

Differences between male and female students were examined next. Both Welch's t-test and Mann-Whitney U-test results were statistically significant for three of the four MSA scales (WE, PG, CD). The MSA scale for EN approached but did not reach statistical significance. These results suggest mean MSA scale scores systematically differ between male and female students. Visual inspection of the distribution plots revealed female

students' mean scores were higher across all MSA scales when compared to their male peers.

Finally, differences between sojourners and non-sojourners were examined. As Table 5 illustrates, none of the Welch's t-tests or Mann-Whitney u-tests indicated there was any difference, for any MSA scale, between students who studied abroad and those who did not.

Discussion and Implications

Research question one asked *to what extent does the MSA appear to be a valid and reliable instrument in the Japanese context?* Taken at face value, the results suggest Japanese students are highly motivated to study abroad, particularly so in terms of how study abroad is connected to their potential world enlightenment or personal growth. Career development is a less-endorsed but still positive motivational factor for studying abroad. Japanese students report being far less motivated to study abroad for the purposes of entertainment. These findings essentially mirror those of Anderson and Lawton (2015).

Upon closer inspection, however, the results also raise key questions. One strength of Rasch analysis is the ability to identify outlying items and respondents. In the case of the MSA, however, the typical measures used for identifying outlying respondents (infit and outfit MNSQ) were of limited use due to such a high percentage of respondents (nearly 20% of the sample) whose scale scores for WE and PG were at or extremely near the maximum possible. I suspect these participants were responding to MSA items in good faith, and I have no reason to suspect they are truly outlier. MSA scores were high across the entire sample for WE and PG.

Research question two asked *in what ways do aspects of study abroad motivation, as measured by the MSA, differ between schools, gender, and those who study abroad versus those who do not study abroad?* The findings of this study illustrate there was significant difference between overall MSA motivational profiles both in terms of the two universities where the study was conducted and between male and female students. This was the case across all MSA scales. These findings do not align with those of Salisbury et al. (2010), who found that interest in study abroad did not differ by gender. However, the profiles of Momiji University and Sakura University possibly explains these differences. Momiji University is a large, national, highly ranked co-educational university. Sakura University is a small, private, low-ranked, Christian, women's liberal arts university. The previous literature (NAFSA, 2008; Salisbury, et al. (2009) suggests student major and/or curricular commitments could explain study abroad participation rates. In Japan, JASSO (n.d.) data

clearly indicates non-STEM majors are participating in study abroad in much higher numbers than their STEM counterparts. It could be that male/female differences in MSA scores is partially attributable to the differences in their respective universities.

There was no significant difference between sojourners and non-sojourners on the MSA scales. What, then, ultimately differentiates between those who sojourn abroad and those who do not? Several factors which are not measured by the MSA or included in this study could explain this difference. Six practical factors identified by Naffziger et al. (2010) were fear of the unknown, curricular challenges, finances, previous commitments, previous travel experiences, and social obligations. Similarly, Salisbury et al. (2009) cites socio-economic status, pre-college capital accumulation, and personality traits such as openness to diversity. In Japan, Oka et al. (2018a) also found openness to experience to be a key factor related to the decision to pursue study abroad.

Alternatively, it is possible participants are not responding to the MSA items not in terms of whether the item motivates *them to pursue study abroad*. Instead, I believe they might be responding in terms of whether the items represent a potential benefit of study abroad-*if they were to participate in study abroad*. If this is possibly the case, it would follow that the benefits of study abroad are mostly agreed upon by Japanese undergraduates, but the intent to pursue study abroad and realize these benefits differs between individuals.

Limitations and Future Research

Several limitations must be acknowledged. First, both the sample size as well as the participant pool coming entirely from one university makes drawing definitive inferences beyond this university difficult. Results should be interpreted with caution. The nature of the survey MSA scales themselves also warrant caution. As previous work by Sponseller (2020) has indicated, there is some reason to believe the MSA scales consist of items which are too easily endorsed. Finally, the data presented here is cross-sectional and was collected at a point in the semester when most students who were going to study abroad the following summer had already enrolled in those programs. It is therefore possible motivation scores were inflated for those who had already decided to study abroad and deflated for those who had decided not to do so or who had been denied entry into their study abroad program(s) of choice.

Future research exploring motivation and study abroad might consider several of the following suggestions. First and foremost, future work exploring Japanese student motivation to study abroad should look to approach the issue with a solid theoretical framework. The international literature on study abroad draws from diverse theoretical

frameworks, and a unified theory of study abroad motivation has yet to be established. Exemplar work here is that of Salisbury et al. (2009, 2010) which draws primarily upon the work Bourdieu (1986) and a student-choice model (Paulsen & St. John, 2002; St. John & Asker, 2001). Future work might build upon this framework while also attempting to integrate either sociocultural theories or educational policies more specifically relevant to Japan. A clear example of a policy which influenced study abroad in Japan is the Ministry of Economy, Trade, and Industry's (METI) push for the development of Japanese young people into global human resources (e.g. *global jinzai*; see Yonezawa, 2014). Furthermore, there is evidence personality theory might explain individual variance in Japanese students' decision to study abroad (Oka et al., 2018a).

Second, a stronger theoretical grounding in future work would allow for the testing of more complex models. One potentially fruitful approach would be to begin by distinguishing study abroad motivation from the actual intent to pursue study abroad. Sanchez et al. (2006) offers a straightforward model in which the intent to study abroad is influenced by study abroad motivations but heavily moderated by financial, familial, and sociocultural barriers. (see Figure 5).
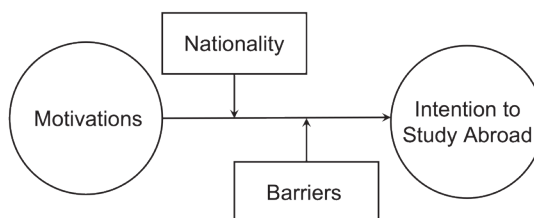
Figure 5　Model exploring motivation to study abroad and intent to study abroad

Including measures of socioeconomic status, at the very least, appears wise. Such work might also include measures of pre-college and in-college social capital accumulation, as in Salisbury et al. (2009).

Third, specific methodological improvements to investigating study abroad are warranted. In general, cross-sectional and purely quantitative approaches to investigating study abroad motivation should be eschewed in favor of longitudinal mixed methods approaches. Additionally, a wider range of scales should be considered. The MSA measures just four motivational aspects connected to study abroad. At the very least, researchers in Japan should explore the validity of Aresi, et al.'s (2017) MMSAS in the Japanese context. The MMSAS consists of 27 items and nine scales, meaning it is of similar length to the MSA yet measures five additional dimensions of study abroad motivation. Of particular importance to study abroad researchers (Allen, 2010; Goldstein &

Kim, 2005) is understanding the role language proficiency or language learning motivation plays when students are determining whether or not to pursue study abroad opportunities. In the case of Japan, work by Oka et al. (2018b) and Churchill and Dufon (2006) indicate foreign language proficiency could present an additional barrier to actively pursuing study abroad. Beyond linguistic proficiency, would-be sojourners' perceptions of self-efficacy in foreign language listening (Kramer & Denison, 2014), speaking (Hicks & McLean, 2014), and intercultural communication (Peterson, et al., 2011; Ferguson, et al., 2017) are also worthy of inclusion when investigating motivation and intent to pursue study abroad opportunities.

Conclusion

Over the last decades, study abroad participation rates have dramatically risen in Japan. Myriad factors appear to influence students' decisions to study abroad (or not). These include financial considerations, gender, academic year of study, and student major. Modern theory on basic psychological needs, however, argues that motivations are multifaceted and directly inform subsequent behaviors (Deci & Ryan, 2012). Developing an instrument capable of reliably measuring aspects of motivation to study abroad is of immense value to future research. Several attempts to develop such an instrument in recent years indicate scholars worldwide agree this is a valuable pursuit (Anderson & Lawton, 2015; Aresi, et al., 2017; Nyapane, et al, 2010; Sanchez, et al., 2006).

The research presented here represents a critical step towards developing such an instrument for use in the context of Japanese undergraduates. This research explored the reliability of a Japanese version of Anderson and Lawton's (2015) Motivation to Study Abroad scale. The alphas and Rasch reliability estimates indicated the Japanese version of this instrument is a reliable measure. Problematically, the scales of World Enlightenment and Personal Growth are almost universally endorsed by Japanese students. In practical terms this means these scales, as they currently exist, are unlikely to adequately separate those who complete them. The absence of such variance almost certainly makes hypothesis testing of the most intriguing types of study-abroad related questions challenging. For example, whether students with different MSA profiles pursue study abroad or benefit from study abroad in different ways will be hard to determine if these scales are so highly endorsed and intercorrelated.

At the institutional level, tests of mean differences revealed significant mean differences between students at Momiji and Sakura University, as well as significant differences between male and female students. The results should be interpreted cautiously given the fundamental differences between the student populations at Momiji

and Sakura University. Perhaps most importantly, there was no significant difference in MSA scale scores between participants bound for a study abroad program and those who were not. Precisely how aspects of motivation to study abroad translate into actually pursuing study abroad remains unelucidated among Japanese undergraduates. Future work should pursue scale development or model testing from a more coherent theoretical framework. A clearer understanding of how study abroad motivational factors influence the intent to study abroad should ultimately benefit educators, program designers, researchers, administrators, policymakers - indeed any group or individual stakeholder with a vested interest in understanding study abroad or the outcomes study abroad experiences might produce.

References

Allen, H. W. (2010). Language-learning motivation during short-term study abroad: An activity theory perspective. *Foreign Language Annals*, *43* (1), 27-49. https://doi.org/10.1111/j.1944-9720.2010.01058.x

Anderson, P. H., Hubbard, A., & Lawton, L. (2015). Student motivation to study abroad and their intercultural development. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *26*, 39-52. https://doi.org/10.36366/frontiers.v26i1.354

Anderson, P. H., & Lawton, L. (2015). The MSA: An instrument for measuring motivation to study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *26* (1), 53-67. https://doi.org/10.36366/frontiers.v26i1.357

Andrich, D. (1978). Scaling attitude itesms constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, *38*, 665-680. https://doi.org/10/dfdd4k

Aresi, G., Alfieri, S., Lanz, M., Marta, E., & Moore, S. (2017). Development and validation of a Multidimensional Motivations to Study Abroad Scale (MMSAS) among European Credit Mobility Students. *International Journal of Intercultural Relations*, *63*, 128-134. https://doi.org/10.1016/j.ijintrel.2017.10.004

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the*

human sciences (3rd ed.). Routledge.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). Greenwood Press.

Churchill, E., & DuFon, M. A. (2006). Evolving threads in study abroad research. In M. A. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 1-27). Multilingual Matters.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416-436). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n21

Dessoff, A. (2006). Who's not going abroad? *International Educator*, *15* (2), 20-27.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Erlbaum.

Elwood, J. A. (2011). Enriching structural models of L2 willingness to communicate: The role of personality, ego permeability, and perceived distance [Unpublished doctoral dissertation]. Temple University.

Ferguson, P., Sponseller, A., & Yamada, A. (2017). Introducing the family reading project. *Transformation in Language Learning*, 129-134.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21* (1), 1095.

Goldstein, S. B., & Kim, R. I. (2005). Predictors of US college students' participation in study abroad programs: A longitudinal study. *International Journal of Intercultural Relations*, *30*, 507-521.

Gnoth J. (1997). Tourism motivation and expectation formation. *Annals of Tourism Research*, *24* (2), 283-304.

Hicks, D., & McLean, S. (2014). *The validation of WTC measurement instruments* [Presentation]. JALT 2014 40th Annual International Conference on Language Teaching and Learning, and Educational Materials Exhibition, Tsukuba.

Hofmeyr, A. S. (2021). Rethinking the concept of global human resources in the Japanese higher education context. *Asia Pacific Journal of Education*, 1-17. https://doi.org/10/gksdt2

Hull C. (1943). *Principles of Behavior*. Appleton-Century-Croft.

JASP Team (2021). JASP (Version 0.14.1) [Computer software].

JASSO. (n.d.). *Nihonjin gakusei ryuugaku joukyouchousa* [Survey of Japanese students studying abroad]. Retrieved September 19, 2021, from https://www.studyinjapan.go.jp/ja/statistics/nippon/index.html

Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. Springer.

Kitsantas, A. (2004). Studying abroad: The role of college students' goals on the development of cross-cultural skills and global understanding. *College Student Journal*, *38* (3), 441-452.

Kramer, B. L., & Denison, C. (2016). *Accurately measuring L2 listening self-efficacy* [Presentation]. JALT 2016 42nd Annual International Conference on Language Teaching and Learning, and Educational Materials Exhibition, Nagoya.

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, *43* (3), 489-493. https://doi.org/10.1016/j.jrp.2008.12.005

Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing:

Person and item contributors to middle category endorsement intent on Likert personality indicators. *Journal of Research in Personality*, *47*(4), 254-262. https://doi.org/10.1016/j.jrp.2013.01.014

Li, M., Olson, J. E., & Frieze, I. H. (2013). Students' study abroad plans: The influence of motivational and personality factors. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *23* (1), 73-89. https://doi.org/10/gkmw44

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3* (2), 103-122.

Linacre, J. M. (2018). Winsteps (4.3.0) [Computer Program]. MESA.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47* (2), 149-174. https://doi.org/10/cssfxp

Naffziger, D. W., Bott, J. P., and Mueller, C. B. (2010). Study abroad: Validating the factor analysis of student choices. *International Business: Research, Teaching and Practice*, *4* (1).

NAFSA: Association of International Educators. (2008). *Strengthening study abroad: Recommendations for effective institutional management*. Report of the Task Force on Institutional Management of Study Abroad.

Nyaupane, G. P., Paris, C. M., & Teye, V. (2011). Study abroad motivations, destination selection and pre-trip attitude formation. *International Journal of Tourism Research*, *13* (3), 205-217. https://doi.org/10.1002/jtr.811

Oka, H., Ishida, Y., & Hong, G. (2018a). Study of factors related to the attitudes toward studying abroad among preclinical/clinical undergraduate dental students at three dental schools in Japan. *Clinical and Experimental Dental Research*, *4* (4), 119-124. https://doi.org/10/gkmw45

Oka, H., Ishida, Y., Hong, G., & Nguyen, P. T. T. (2018b). Perceptions of dental students in Japanese national universities about studying abroad. *European Journal of Dental Education*, *22* (1), e1-e6. https://doi.org/10.1111/eje.12212

Paulsen, M. B., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *The Journal of Higher Education*, *73* (2), 189-239.

Peterson, J. C., Milstein, T., Chen, Y.-W., & Nakazawa, M. (2011). Self-efficacy in intercultural communication: The development and validation of a sojourners' scale. *Journal of International and Intercultural Communication*, *4* (4), 290-309. https://doi.org/10.1080/17513057.2011.602476

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Salisbury, M. H., Paulsen, M. B., & Pascarella, E. T. (2010). To see the world or stay at home: Applying an integrated student choice model to explore the gender gap in the intent to study abroad. *Research in Higher Education*, *51*(7), 615-640. https://doi.org/10.1007/s11162-010-9171-6

Salisbury, M. H., Umbach, P. D., Paulsen, M. B., & Pascarella, E. T. (2009). Going global: Understanding the choice process of the intent to study abroad. *Research in Higher Education*, *50* (2), 119-143. https://doi.org/10.1007/s11162-008-9111-x

Sanchez, C. M., Fornerino, M., & Zhang, M. (2006). Motivations and the intent to study abroad among U.S., French, and Chinese students. *Journal of Teaching in International Business*, *18* (1), 27-52. https://doi.org/10/b689t3

Sponseller, A. C. (2020). Initial validity evidence for a Japanese version of the Motivation to Study Abroad Scale (MSA). *OnCUE Journal Special Issue*, *1*, 185-199.

St. John, E. P., & Asker, E. H. (2001). The role of finances in student choice: A review of theory and research. In M. B. Paulsen & J. C. Smart (Eds.), *The finance of higher education: Theory, research, policy, & practice* (pp. 419-438). Agathon Press.

Vroom, V. H. (1964). *Work and motivation*. Wiley.

Wang, P. Y. (2004). Teacher as leader and student as follower: The implementation of expectancy theory in the English classes of Taiwanese college. *Journal of American Academy of Business*, *5* (1/2): 418-422.

Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, *6* (1), 80-99.

Wiers-Jenssen, J. (2003). Norwegian students abroad: Experiences of students from a linguistically and geographically peripheral European country. *Studies in Higher Education*, *28* (4), 391-411. https://doi.org/10.1080/0307507032000122251

Williams, E. A. (2020). The role of English and images of 'the international' in the lives of young Japanese male sojourners. *Doshisha Women's University Academic Research Annual Report*, *71*, 29-42.

Wolfe, E. W., & Smith Jr, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part I-instrument development tools. *Journal of Applied Measurement*, *8* (1), 97-123.

Yonezawa, A. (2014). Japan's challenge of fostering "global human resources": Policy debates and practices. *Japan Labor Review*, *11* (2), 37-52.

APPENDIX
The Motivation to Study Abroad Scale

How simportant is each of the following in motivating you to participate in a study abroad program?
留学参加動機として、以下の事柄はあなたにとってどのくらい重要ですか。

Response scale:

1 – not at all important; 全く重要でない

2 – not important; 重要でない

3 – not very important; あまり重要でない

4 – a little important; 少し重要である

5 – important; 重要である

6 – very important; 非常に重要である

World Enlightenment

MSA 4 Become acquainted with people different from me（自分から人と知り合いになる[a]）

MSA 8 Increase my understanding of the world（世界をより理解する）

MSA 11 Expand my world view（世界を見る目を広げる）

MSA 15 Enhance my understanding of global affairs and events（世界的な出来事や事柄をより理解する）

MSA 18 Interact with people from other countries（他国の人々と交流する）

MSA 21 Better understand different cultures（多文化を理解する）

MSA 23 Learn about the world（世界を学ぶ）

Personal Growth

MSA 1 Gain maturity（成長すること）

MSA 5 Better understand myself（自分をよりよく知る）

MSA 6 Increase my self-confidence（自信を高める）

MSA 14 Grow as a person（人として成長する）

MSA 20 Become more independent（より自立する）

MSA 22 Learn to stand on my own two feet（自立する事を学ぶ）

Career Development

MSA 3 Prepare for my career（自分のキャリアの準備）

MSA 7 Enhance my employment prospects（就職の将来性を高める）

MSA 9 Gain career skills（職業的スキルを上げる）

MSA 17 Gain in-depth knowledge in my chosen field（自分の専攻における知識を増やす）

MSA 19 Build my resume（履歴書に書く）

Entertainment

MSA 2　Experience the local nightlife（clubs, bars, etc.）（その地域のナイトライフ（クラブ、バーなど）の体験）

MSA 10 Have a romantic encounter（異性との出会い）

MSA 12 Go out drinking（飲みにいく）

MSA 13 Make my friends a little envious of me（友だちに少し羨ましいと思わせる）

MSA 16 Do some serious partying（パーティーをする）

Regarding citation:

When using the English items, cite the original authors:

Anderson, P. H., & Lawton, L. (2015). The MSA: An instrument for measuring motivation to study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *26* (1), 53-67. https://doi.org/10.36366/frontiers.v26i1.357

When using the Japanese items, cite the following:

Sponseller, A. C. (2020). Initial validity evidence for a Japanese version of the Motivation to Study Abroad Scale (MSA). *OnCUE Journal Special Issue*, *1*, 185-199.

---

[a] As noted by one of the reviewers, the Japanese translation for MSA 4 could be problematic. The translation above was used in the research presented in this manuscript, thus it is presented here.