

Investigating Stereotype Threat in Foreign Language Acquisition: A Preliminary Study¹

Matthew Nall², Miyagi University, Sendai, Miyagi Prefecture, Japan

Abstract

Various research studies in the field of psychology have shown that the phenomenon of stereotype threat is real and has a consistent and measurable negative effect on individuals' performance on cognitively demanding tasks in a variety of areas. Currently, the relationship between stereotype threat and second language acquisition has not been explored in empirical research. This preliminary study aims to fill this gap by identifying stereotypes in regard to English and English as a foreign language (EFL) education in Japan, and testing whether stereotype threat has a measurable and consistent impact on English acquisition by Japanese university students. Results show that stereotype threat plays a role in English acquisition in Japan. As the current study is only preliminary in nature, further research is warranted. This article concludes with a discussion of limitations and implications, and gives some guidance for further research in this new area of inquiry.

Resumen

Varios estudios de investigación en el campo de la psicología han demostrado que el fenómeno de la amenaza del estereotipo es real y tiene un efecto negativo constante y medible en el desempeño en tareas cognitivamente exigentes en una variedad de áreas. Actualmente, la relación entre la amenaza de los estereotipos y la adquisición de un segundo idioma no se ha explorado mediante investigación empírica. Este estudio preliminar tiene como objetivo llenar este vacío identificando estereotipos con respecto al aprendizaje del inglés como lengua extranjera (ILE) en Japón. Busca saber si la amenaza de los estereotipos tiene un impacto medible y constante en la adquisición del inglés por parte de estudiantes universitarios japoneses. Los resultados muestran que la amenaza del estereotipo juega un papel en la adquisición del inglés en Japón. Dado que el estudio actual es solo de naturaleza preliminar, se justifica una mayor investigación. Este artículo concluye con una discusión de las limitaciones e implicaciones, y brinda una guía para futuras investigaciones en esta nueva área de investigación.

Introduction

Researchers in Second Language Acquisition (SLA) strive to understand how effective language learning takes place and how to incorporate that body of knowledge into classroom practice (Barkhuizen, 2018). While most people have little difficulty acquiring their first language (L1), there is a wide spectrum of learner success in the acquisition of a second language (L2) (Dörnyei, 2005; Lightbown & Spada, 2013; Sasayama, 2018). Individual characteristics of the learner can account for much of this variation. These characteristics include personality, attitude, motivation, beliefs, and age, among others (Dörnyei, 2005; Lightbown & Spada, 2013). Emotional responses to language learning have also earned an important place in second language research and theory. Krashen's (1982) Monitor Model includes the discussion of the affective filter—the idea that a learner's acquisition of a second language is constrained when they experience negative emotions about the language. Affect is understood to be "emotionally relevant characteristics of the individual that influence how he/she will respond in any situation" (Gardener & MacIntyre, 1993, p. 1). Emotional responses can manifest as anxiety about learning or using the language, low self-esteem or self-confidence, and low levels of willingness to communicate (Thurman, 2018). While individual characteristics and emotional responses are important in language learning scenarios, establishing a framework that accounts for the role of stereotype threat in the context of SLA will contribute to a fuller understanding of how languages are learned. The aim of the present study is to determine whether the phenomenon of stereotype threat can account for variations in second language learner success, and to establish stereotype threat as a research niche within fields related to language learning and teaching.

Language teachers and researchers know that although many students begin the journey of L2 study, few achieve high levels of L2 proficiency. This is especially true in Japan, where English education is a compulsory subject in secondary education (Kayi-Aydar, 2018). It is important for language teachers and researchers to overcome the barriers that prevent students from realizing their full potential. Individual characteristics and emotional responses of the learner can account for much of this variation. However, there may be some sociocultural and psychological barriers that have yet to be fully explored in this area. Concerning this, research from the field of psychology that identifies, analyzes, and theorizes the phenomenon of stereotype

¹This is a refereed article. Received: 9 April, 2021. Accepted: 14 June, 2021.

² matthewgarynall@gmail.com, 0000-0002-4348-1954

threat has thrown some light on the problem (Ackerman-Barger et al., 2016; Ambady, et al., 2001; Croizet & Claire, 1998; Shih et al., 1999; Steele, 2010). Stereotype threat is a well-founded theory in the field of psychology (Steele, 2010). It is a socioculturally-bound psychological barrier that prevents members of a group, who suffer from an imposed negative stereotype in some domain, from performing at their best or attaining their goals in that domain (Vedantham, 2015). The present investigation examines the relationship between stereotype threat and second language acquisition in order to understand more about how foreign languages are learned, and to determine whether more study is warranted in this new and promising area of inquiry within the field of SLA.

Literature Review

Identifying Stereotype Threat

Across various academic fields, the concept of identity has been an intriguing topic of research (Block, 2007; Hermans & Hermans-Konopka, 2010; Norton, 2017; Pavlenko & Blackledge, 2004; Peirce, 1995; Toohey, 2001). To contribute to this ongoing academic discussion, social-psychologist Claude Steele (2010) defined identity contingencies. These are "the things that you have to deal with in a situation because you have a social identity" (p. 3). Contingencies are variables such as one's age, gender, race, sexuality, political affiliation, and health (Steel, 2010). While the concept of identity remains an important topic of contemporary research and theory, what is known, is that an individual's identity plays a significant role in determining social situations and how individuals experience and respond to those situations.

Central to Steele's (2010) thesis is the phenomenon of stereotype threat. "Stereotype threat is being at risk of confirming, as a self-characteristic, a negative stereotype about one's group" (Steel & Aronson, 1995, p. 797). Shankar Vedantham illustrates this concept. "Let's say that you think people have a certain stereotype about you. There's a part of you that's afraid that your actions and behavior will prove the stereotype true" (Vedantham, 2015, p. 2). Research has shown that stereotype threat is a psychological phenomenon that negatively impacts one's performance on a cognitively demanding task when there is a negative stereotype about the individual in that domain (Aronson et al., 1998; Schmader, 2002; Spencer et al., 1999; Steele, 1997; Steele, 2010). As a recognized phenomenon in the field of psychology, researchers from other academic disciplines should be aware of stereotype threat and the negative affects it can have on individuals.

Examining Stereotype Threat in Various Domains

As the psychological phenomenon of stereotype threat gains traction in academia, researchers have increased their efforts to explore and understand stereotype threat in different scenarios. In an influential study on stereotype threat, Shih et al. (1999) investigated the performances of Asian-American women on a math test. Immediately prior to taking the test, participants in the control and experimental groups completed one of two identity-salience surveys. These surveys were designed to activate the part of the participants' personality that identifies with either being female, or being Asian. This method was chosen because in North America, there is a positive stereotype—*Asians are good at math*—but there is also a negative stereotype—*women are not good at math*. The authors wanted to pit these two stereotypes against each other to see whether stereotype threat would affect the participants' test performance. The hypothesis was that due to the influence of stereotype threat, the Asian-salient group would outperform the female-salient group on the test. Results showed that the Asian-identity-salient participants did remarkably better than the female-identity-salient group, even though the research method included a control for participant ability. It was determined that this difference was due to the influence of stereotype threat. The authors concluded that "performance is malleable and sensitive to situational and psychological cues" (p. 389). This study has served as a basis and model for continued research on stereotype threat.

Some studies have looked specifically at stereotypes and stereotype threat as they are evident along gender lines. Pavlova et al. (2014) explored gender-made stereotypes and stereotype threat using a story sequencing task. The authors showed that when given a negative oral prompt before a task—*women generally are worse at this than men*—women were disproportionately affected by this when compared to males. Brain imaging showed that in women, areas of the brain that are responsible for processing negative emotions were much more active during the task than they were in men's brains and this inhibited their overall performance. The women's underperformance on the task was attributed to the additional cognitive processing going on in the women's brains. In the same vein, a study by Brown and Pinel (2003) showed that women's performances varied according to how conscious they were of stereotypes

Children have also been an important focus in stereotype threat research. Ambady et al. (2001) wanted to know when children become aware of stereotypes and therefore become susceptible to the negative effects of stereotype threat. Findings showed that children as young as 5-7 years old were aware of stereotypes, and could be influenced by stereotype threat. Similar results were confirmed by Muzzatti and Agnoli (2007). Another study looked at the relationship between mothers' beliefs regarding stereotypes, and their children's susceptibility to stereotype threat. Tomasetto et al. (2011) showed that mothers' endorsement of stereotypes can have an important impact on their children's susceptibility to stereotype threat. This study considered the performances of daughters on a math test. In this experiment, parents first completed a survey that identified their endorsement of stereotypes concerning girls and math. Then their daughters completed a math test along with participants whose parents showed no belief in these stereotypes. Results showed that the daughters of the parents endorsing a negative stereotype did significantly worse on the test than participants whose parents did not express belief in them.

Having established stereotype threat as a phenomenon that can inhibit intellectual performance, it was necessary to determine the effects of stereotype threat over time. In an investigation of stereotype threat among minority health care professionals, Ackerman-Barger et al. (2016) identified short and long-term effects of stereotype threat. Short-term effects included hypervigilance, impaired working memory and self-regulation, and inhibited intellectual performance. Long-term effects included disengagement from the healthcare domain, and dis-identification. In an earlier study, Rothgerber and Wolsiefer (2014) looked at the role of stereotype threat among young male and female chess players and found similar long-term effects. Female players affected by stereotype threat were the least likely to continue playing in future tournaments (i.e., disengagement from the domain). In an effort to preemptively identify at-risk individuals and implement early interventions, Picho and Brown (2011) designed a survey that was intended to measure the level of participants' susceptibility to stereotype threat. The authors claimed that their survey could be used to identify at-risk individuals, as well as aid in the development of interventions that would help them respond in a positive and productive manner.

In sum, research from the field of psychology has shown that stereotype threat is a genuine phenomenon that has a consistent and quantifiable negative effect on individuals' performances on a variety of cognitively demanding tasks in many domains. Subtle activation of a part of a person's identity prior to completing a task can affect their performance on that task (Aronson et al., 1998; Shih et al., 1999; Steel, 2010). As indicated above, children are susceptible to stereotype threat from a very young age (Ambady et al., 2001; Muzzatti & Angoli, 2007; Tomasetto et al., 2011) and women can be disproportionately affected by it (Pavlova et al., 2014). There are many detrimental short-term and long-term effects of stereotype threat (Ackerman-Barger et al., 2016; Rothgerber & Wolsiefer, 2014). It is clear then, that the phenomenon of stereotype threat is an important issue that deserves to be the subject of ample research in various fields.

Stereotype Threat and Second Language Acquisition

Given the well-established body of knowledge concerning stereotype threat in the field of psychology, there is a need for researchers across various disciplines to give stereotype threat consideration in empirical research studies within those academic fields. Recently, a few research studies have examined gender stereotypes in foreign language learning (Chaffee et al., 2020; Kutuk, 2019). But there remains a dearth of empirical research which examines the relationship between stereotype threat and second language acquisition in broader sociocultural contexts. The current research study aims to begin to fill this gap in understanding. As stereotypes are culturally bound, this research was carried out specifically in the context of English language learning in Japan.

Within fields of second language learning and teaching there are ample opportunities to conduct studies on stereotype threat. Researchers in SLA can begin to understand the relationship between stereotype threat and second language acquisition by conducting research in specific sociocultural contexts. Studies in this area have the advantage of being able to draw upon the established epistemologies and methodologies that have been developed and tested in previous studies conducted in psychology. For example, oral and written prompts have been shown to elicit stereotype threat in several studies as outlined above (e.g., Ambady et al., 2001; Pavlova et al., 2014; Shih et al., 1999; Tomasetto et al., 2011). These methodologies can be replicated in second language learning scenarios in order to determine whether stereotype threat is in fact an issue that can affect language learner success. Evidence in this area can help account for variation in learner success. If it can be established that stereotype threat can inhibit the language learning process,

future studies might identify individuals that are most susceptible to stereotype threat. These studies may be based on survey methods developed by Picho and Brown (2011), and researchers may implement early interventions (e.g., Ackerman-Barger et al., 2016) so that individuals may mitigate the pitfalls of stereotype threat and find greater success in language learning.

The present research aims to determine whether stereotype threat does in fact play a role in second language acquisition. Given the abundance of research in psychology concerning stereotype threat, it is reasonable to hypothesize that stereotype threat plays some role in the second language learning process. However, this has not yet been evidenced in empirical research. This study aims to investigate this area, and if warranted, establish stereotype threat as a research niche within the fields of second language learning and teaching for further inquiry.

Methods

Research Questions

The primary research questions (RQs) addressed in the study are as follows:

1. *What stereotypes do Japanese university students endorse concerning English?*
2. *Does stereotype threat have a quantifiable effect on English performance of Japanese university students?*

Research Design

The current study uses a quasi-experimental design. The research plan (Figure 1) was carried out according to university research ethics guidelines. An ethics review was completed before the collection of any data. Participants gave written informed consent, and were not coerced in any way into participating. Survey data in the study were collected anonymously.

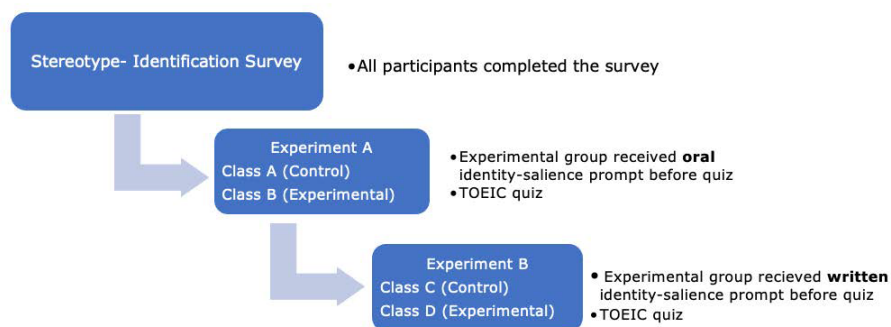


Figure 1: Research plan

In prior studies on stereotype threat, identity-salience activities given before cognitively demanding tasks took a variety of forms. These have been shown to elicit a stereotype threat response. In the present study, the researcher tested the effectiveness of two different forms of identity-salience activities that were used in prior studies—oral prompts (e.g., Pavlova et al., 2014) and written prompts (e.g., Ambady et al., 2001; Muzzutti & Agnoli, 2007; Picho & Brown, 2011; Shih et al., 1999; Tomasetto et al., 2011). In order to test these individually, the research was divided into two separate experiments—Experiment A and Experiment B. All participants were asked to complete a practice Test of English for International Communication (TOEIC) quiz and the primary dependent variable was accuracy. Before the quiz, Experiment A required participants in an experimental group to listen to an identity-salience presentation (i.e., an oral prompt) about the under-performance of Japanese test takers when compared to their closest geographic neighbors. Experiment B involved the use of an identity-salience survey (i.e., a written prompt), based on the survey methods used in Shih et al. (1999), which required participants to answer questions about the part of their personality that identified with either being female or Asian. In the present study, the identity-salience survey was designed to activate the part of the participants' personalities that identified with being Japanese in order to elicit a stereotype threat response. In both Experiment A and Experiment B, control groups took a practice TOEIC quiz. Before the TOEIC quiz, the experimental groups engaged in an identity-salience activity (oral prompt for Experiment A; a written prompt for Experiment B), then took the TOEIC quiz. The same TOEIC quiz was utilized in Experiment A and Experiment B.

Participants

The participants ($n = 142$) were first-year students at a public university in northern Japan. All participants spoke Japanese as their first language. Participants had similar levels of English proficiency as evidenced by their actual TOEIC scores from a prior semester. They all had studied English formally from their first year of junior high school through their first year of university courses. Participants were divided into groups: Class A ($n = 33$), Class B ($n=37$), Class C ($n=36$), and Class D ($n=36$). Due to educational constraints, the participants could not be randomly assigned to the control or experimental groups. Instead, the researcher used class sections (Class A, B, C, D) of an existing English course that he taught. None of the participants was an English major.

Research Procedure

The researcher addressed RQ 1 by administering an anonymous survey (Table 1) to all participants in the study. Before any data collection took place, they were informed that the current research study was about the relationship between stereotypes and language learning, and that they were not coerced in any way into participating in the study. Having obtained written informed consent, the survey was conducted in Japanese to ensure valid results and included eight questions ranked on a 5-point Likert scale (strongly agree – agree – neutral – disagree – strongly disagree). The survey questions were designed by the researcher based on stereotype categories (e.g., gender or race) identified in prior research studies (e.g., Ambady et al., 2001; Muzzutti & Agnoli, 2007; Picho & Brown, 2011; Shih et al., 1999; Tomasetto et al., 2011). The aim of the survey was to identify existing stereotypes about English that are held by Japanese university students. In order to support instrument reliability, the survey was piloted to make sure that questions were not ambiguous, and that participants did not become fatigued, misinterpret questions, or guess the expected answers. Survey questions also included a measure of internal consistency and reliability by asking stereotype-related items in various ways (e.g., Q3 and Q4 in Table 1).

Survey data were analyzed (see Table 3 below), and the researcher found that there were sufficient data to support the existence of specific English-related stereotypes (please refer to subheading *The Stereotype threat identification survey results* below for discussion of survey results). Stereotypes held by the majority of participants were identified, so the researcher determined that it would be appropriate to move on to the experimental portion of the study. This consisted of two experiments (Experiments A and B) in order to test the reliability of two different identity-salience activities (oral and written prompts) that had been established in prior research studies.

-
- Q1. In general, Japanese people are bad at English.
 Q2. Japanese people that are good at English are cool.
 Q3. Japanese boys are better at English than girls.
 Q4. Japanese girls are better at English than boys.
 Q5. I like studying English.
 Q6. Compared to my peers, I am better at English.
 Q7. English is important for my future.
 Q8. I want to use English at my future job.
-

Note. Actual surveys were conducted in Japanese.

Table 1: *Stereotype-identification survey questions*

Experiment A

Based on the methodologies of Pavlova et al. (2014), the researcher wanted to measure the effect of an oral prompt as a catalyst for stereotype threat. Participants were divided into control (Class A) and experimental (Class B) groups. Participants had similar English abilities, with the control group having an actual mean TOEIC score from a previous semester of 411 ($SD = 105$), and the experimental group having a mean score of 397 ($SD = 110$). During the last 15 minutes of a regular class, control group participants were asked to complete a practice TOEIC quiz. The quiz consisted of eight fill-in-the-blank multiple-choice questions and four multiple-choice reading comprehension questions. Participants were told that they did not have to complete the quiz and were free to leave if they wanted to do so, as the quiz would have no effect on their grades. After handing out the quiz, the researcher left the classroom. Quizzes were timed and collected by a designated volunteer and returned in a sealed envelope to the researcher for analysis.

The experimental group was asked to complete the same TOEIC quiz as the control group. However, immediately before taking the quiz, participants listened to an identity-salience presentation (i.e., oral

prompt). For this, participants were presented with the 2018 TOEFL iBT mean scores listed by country (Educational Testing Service, 2018) (Figure 2), as part of a presentation given by the researcher. Since the data for the TOEFL iBT test was organized by country and readily available, the researcher included this test data in the experiment. During the mini-presentation, the participants' attention was drawn to the fact that the average TOEFL iBT score of test-takers in Japan was below those of most other Asian countries. According to the data, Japan has one of the lowest average scores in Asia (Japan's average score = 71), while Japan's closest geographic neighbors, such as China (80), South Korea (84), and Taiwan (82), outpaced Japan on the test in nearly every aspect. Quizzes were then distributed to the class, and the researcher then left the classroom. Quizzes were timed and collected by a designated volunteer and returned in a sealed envelope to the researcher for analysis.

ASIA					
Afghanistan	15	17	21	19	72
Azerbaijan	18	20	21	20	80
Bangladesh	21	22	22	22	87
Bhutan	*	*	*	*	*
Brunei Darussalam	*	*	*	*	*
Cambodia	16	18	19	19	72
China	21	19	19	20	80
Christmas Island	*	*	*	*	*
Cocos (Keeling) Islands	*	*	*	*	*
Hong Kong	21	23	22	23	89
India	23	24	24	24	95
Indonesia	21	22	21	22	86
Japan	18	18	17	18	71
Kazakhstan	19	21	21	20	82
Korea, Democratic People's Republic of	21	21	20	21	83
Korea, Republic of	22	21	20	21	84
Kyrgyzstan	17	19	20	19	74
Lao, People's Democratic Republic	13	16	18	17	64
Macao	20	21	21	21	83
Malaysia	22	23	22	23	90
Maldives	*	*	*	*	*
Mongolia	18	20	20	19	77
Myanmar	19	20	21	21	81
Nepal	20	21	21	22	84
Pakistan	22	23	24	23	92
Philippines	21	22	23	22	88
Singapore	24	25	24	24	98
Sri Lanka	20	21	21	21	84
Taiwan	21	21	20	20	82
Tajikistan	15	17	19	18	69
Thailand	19	20	19	20	78
Timor-Leste	*	*	*	*	*
Turkmenistan	17	19	20	19	75
Uzbekistan	20	21	22	20	83
Vietnam	20	21	20	22	83

Figure 2: 2018 TOEFL iBT results listed by country

Figure 2 shows scores for reading, writing, listening, and speaking sections. Overall mean test scores are found in the far-right column (Educational Testing Service, 2018).

Experiment B

In this experiment, the researcher wanted to test the effect of an identity-salience survey (i.e., a written prompt) as a catalyst for stereotype threat. Experiment B was conducted similarly to Experiment A, however different participants were chosen in this portion of the study. Participants in the control group (Class C) took only the TOEIC quiz. The experimental group (Class D) took the same quiz after taking part in a different identity-salience activity (Table 2), which was based on the identity-salience survey designs used by Shih et al. (1999). The two groups had similar mean TOEIC scores in the previous semester, with Class C having a mean score of 451 (SD = 117), and Class D with a mean score of 479 (SD = 106). All survey questions were based on survey methods utilized by Shih et al. (1999) and were designed by the current researcher to activate the part of the participant's personality that identifies with being Japanese. The survey was piloted for reliability and checked for ambiguity and vagueness. Sample questions can be seen in Table 2. Identity-salience surveys in Experiment B were conducted in Japanese, and consisted of 10 yes/no questions. TOEIC quiz data was analyzed for both groups.

-
1. Were you born in Japan?
 2. Were both of your parents born in Japan?
 3. When you communicate with your family members, do you speak only Japanese?
 4. Have you lived only in Japan for your entire life?
-

Note. Actual surveys were conducted in Japanese.

Table 2: Sample Japanese-Identity-Salience survey questions

Results and Discussion

The Stereotype-Identification Survey Results

The stereotype-identification survey (Table 1 above) was administered to participants prior to beginning the experimental portion of the study. Results of the Likert-style survey questions can be seen in Table 3. Regarding RQ 1, the data confirms the existence of a negative stereotype concerning the English ability of Japanese people that is endorsed by a large number of participants. By examining the responses of survey question 1 (Q1 in Table 3) — *Japanese people are bad at English*—it is apparent that a negative stereotype exists, as most participants either agreed or strongly agreed with this statement. While many Japanese people endorse a negative stereotype concerning English abilities of Japanese people, results of survey question 2 (Q2)—*Japanese people that are good at English are cool*—indicate that there is admiration for Japanese people who achieve high levels of English proficiency.

	Q1	Q2	Q3	Q4
Strongly agree	41	80	2	1
Agree	50	16	4	9
Neutral	8	2	78	81
Disagree	1	2	12	7
Strongly disagree	0	0	4	2

Note. Survey Questions can be found in Table 1. Surveys were conducted in Japanese.

Table 3: *Results of the Stereotype-Identification Survey Likert-scale questions*

In addition, the stereotype about Japanese people's English abilities, gender stereotypes were also addressed in Q3 (*Japanese boys are better at English than girls*) and Q4 (*Japanese girls are better at English than boys*). These were included in the survey because in many of the stereotypes addressed in prior research studies were divided along gender lines (Ambady et al., 2001; Muzzatti & Agnoli, 2007; Pavlova et al., 2014; Tomasetto et al., 2011). The data suggest that there is no gender-related stereotype in the SLA context in Japan, as participants chose the neutral option in response to the Stereotype-identity survey Q3 and Q4. What is not clear is whether stereotype threat is a factor which influences individuals' performances on English-related tasks. The current researcher determined that the experimental portion of the study was worth undertaking in order to address RQ 2, as a negative stereotype exists in Japan and is endorsed by many, and to explore whether stereotype threat has a quantifiable and consistent effect on English performance.

Experiment A Results

Experiment A TOEIC quiz data (Table 4) were analyzed, and it was found that the scores for the experimental group (Class B), were lower than the scores of the control group (Class A). This relationship was predicted in the primary hypothesis of the current study, and is consistent with the results of prior research studies (Ambady et al., 2001; Pavlova et al., 2014; Tomasetto et al., 2011; Shih et al., 1999).

	Mean score	SD	Mean grade
Experiment A			
Class A (control)	8.5	2.49	53%
Class B (experimental)	8.3	2.44	52%
Experiment B			
Class C (control)	7.3	3.11	46%
Class D (experimental)	6.7	2.35	42%

Table 4: TOEIC quiz results

A *t*-test was conducted for Experiment A quiz data to determine statistical significance of the results. Results were considered to be statistically significant if $p < .05$. Statistically significant results would give reason to believe that stereotype threat can inhibit second language performance. The Experiment A result of the *t*-test was $p = 0.37$. The *p* value is not less than $p = .05$, and therefore the results of Experiment A were not statically significant. Given this result, the researcher cannot say here that stereotype threat had a negative impact on participant performance in this experiment. This was unexpected, as similar studies have yielded statistically significant differences between control and experimental groups (Ambady et al., 2001; Pavlova et al., 2014; Tomasetto et al., 2011; Shih et al., 1999). This finding requires further consideration, which is discussed below.

Experiment B Results

Like Experiment A, the researcher analyzed Experiment B quiz data (Table 4 above). The mean scores and grades for the experimental group (Class D) were lower than the scores and grades of the control group (Class C). This relationship was consistent with the Experiment A results, and with findings of previous studies (Ambady et al., 2001; Pavlova et al., 2014; Tomasetto et al., 2011; Shih et al., 1999). A t-test was conducted, and the results were determined to not be statistically significant because $p = 0.15$, which is not less than 0.05. Once again, the trend of the control group outperforming the experimental group seems to be consistent with previous studies however, the fact that the difference was not statistically significant is again puzzling, and worthy of further inquiry and consideration.

Further Analysis of Experimental Data

Findings for both Experiment A and Experiment B were determined to be not statistically significant. This result was unexpected, as many of the studies in the literature review above (which were not related to SLA or EFL) concluded with statistically significant differences between the control and experimental groups. Statistically significant results would have given credence to the main hypothesis—stereotype threat has a consistent and measurable impact on participants' performance on an English test in the SLA context in Japan. Confirmation of the main hypothesis of the present study was not borne out by the t-test results. Therefore, the researcher returned to a review of prior research studies to re-evaluate the current experimental parameters, and to look for clues that might explain the experimental results.

Prior research studies concerning stereotype threat concluded with some similar findings. For example, a linear contrast analysis conducted by Shih et al. (1999) found that participants under stereotype threat consistently did worse than their control group counterparts. This relationship was found to remain true in Experiment A and Experiment B results of the current study. Going further, in a study investigating ways that the negative effects of stereotype threat can be mitigated, Picho and Brown (2011) found that in order for stereotype threat to have an effect on an individual's performance, there are two prerequisites. First, the individual must believe that the stereotype exists. In the present study, a negative stereotype was found to exist in accordance with the results of RQ 1 and the stereotype-identification survey (Table 3 above). Secondly, the individual has to be invested in the domain. That is to say the participant has desire to do well on the task. This requirement captured the present researcher's attention. This is a measure of the participant's investment or motivation in the domain. For example, if the instrument in a given study is a math test and the primary dependent variable is accuracy, the participant must desire to do well on the math test in order for stereotype threat to come into play. In the present study, the research method did not include a control for participant investment or motivation in English language acquisition. Participants had enrolled in the English classes in order to fulfill graduation requirements, so they had wide variation in levels of investment and motivation. This was a weakness in the research design. Because of this, the researcher analyzed the Experiment A and Experiment B TOEIC quiz data by score range in order to more appropriately address RQ 2. According to Picho and Brown's (2011) claim about prerequisites, individuals who score in the highest percentiles are likely to be highly invested in the domain. Therefore, they are the participants who are most susceptible to stereotype threat.

The Experiment A TOEIC quiz scores (Table 3 above) and Experiment B TOEIC quiz scores (Table 4 above) were broken down according to score range. For the Experiment A experimental group (Class B), the number of participants who scored in the top range (10 or more correct answers) was somewhat lower than the number of participants that scored in the top of the range in the control group (Class A). In Experiment A, only 30% of participants in the experimental group scored in the top range, while nearly 39% of all participants in the control group did so. This data set shows a dramatic drop in the number of participants that scored in the top of the range. Accordingly, there was a slight increase in the number of participants that scored in the mid-range in the experimental group, when compared to the control group, and there was another significant increase in the number of participants in the experimental group that scored in the low-range, when compared to the control group. It appears that when quiz results are separated by range, there is an apparent negative impact in the top-percentile scorers when participants operate in the context of stereotype threat. This result resembles the findings of previous studies on stereotype threat (Ambady et al., 2001; Pavlova et al., 2014; Tomasetto et al., 2011; Shih et al., 1999). Concerning RQ 2, it can be said that stereotype threat does in fact matter for many of those participants who are invested in the domain.

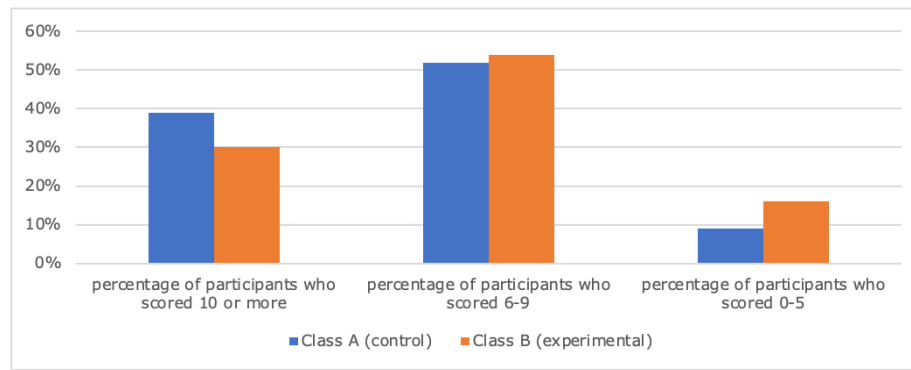


Figure 3: Experiment A-TOEIC quiz results separated by range

Experiment B TOEIC quiz results separated by range can be seen in Figure 5. Just like the results of Experiment A, Experiment B participants in the experimental group with scores in the top range were significantly fewer, dropping from 25% of all participants in the control group to 14% of all participants in the experimental group. This is an impressive drop in the number of participants who scored in the top range. This fact falls in line with the results of Experiment A data separated by range, and gives further support for the main hypothesis and the idea that stereotype threat has a negative impact on language acquisition. When we observe the participants' scores in the mid-range in Experiment B, we see a sharp increase in the number of scores in the experimental group when compared to the control group, and a slight decrease in scores in the experimental group in the low range as compared to the control group. This differs slightly from the results of Experiment A, and deserves further study.

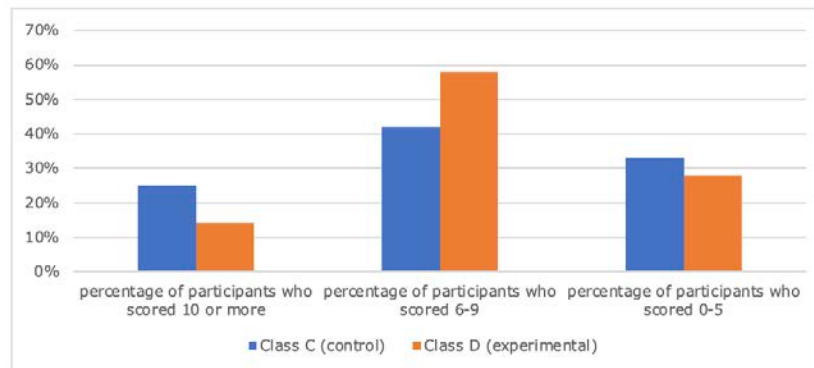


Figure 4: Experiment B TOEIC quiz results separated by range

Additionally, Experiment A groups (Class A and B) had mean TOEIC scores from the previous semester of 411 and 397 respectively. While Experiment B groups (Class C and D) had average TOEIC scores of 451 and 479. On the average, the Experiment B participants had a higher English language proficiency according to their mean TOEIC scores than did Experiment A participants, evidenced by their higher average actual TOEIC scores. When TOEIC quiz data was separated by range, Experiment A saw a 9% drop in the number of participants who scored in the top range (39% of control group participants scored in the top range, while only 30% of the experimental did). Experiment B saw a more dramatic 11% drop in the number of participants that scored in the top end of the range, falling from 25% of participants in the control group to 14% of participants in the experimental group. The significant drop in the number of participants who scored in the top of the range falls shows support for Picho and Brown's (2011) prerequisite concerning motivation and investment in the domain. This study showed evidence that individuals who are more invested in the domain (i.e., they care more about whether they do well in the domain than their other, less invested counterparts) are more likely to be negatively affected by stereotype threat. The present study concluded with similar results, as the number of participants who scored in the top range dropped in the experimental group when compared to the control group. This lends support to Pico and Brown's (2011) theory that participants who are most invested in the domain suffer the most when performing in a context of stereotype threat.

Conclusions

This study is preliminary in nature because the findings were not statistically significant. However, a number of interesting conclusions can be drawn from the data which can inform and guide future research studies, as this is a first foray into a new and exciting area of research that examines the role of stereotype threat in SLA.

Regarding RQ 1, it can be said that in Japan, there is a negative stereotype concerning Japanese people's general English abilities. According to the stereotype-identification survey (Table 2; Table 3), the vast majority of participants agreed with the statement "In general, Japanese people are bad at English." It can also be said that there is no particular gender stereotype regarding English in Japan. Concerning RQ 2, there is some evidence to support the hypothesis that stereotype threat has a negative impact on the performances of people who are invested in English. This relationship deserves more attention in empirical research, but in the present study participants who were more invested in English were more likely to be negatively affected by stereotype threat. This inference was weak, as results were not statistically significant overall. This may have been due to weaknesses in the research design. It is expected that future studies will control for participant investment in English education, and perhaps conclude with more robust results.

Considering the identity-salience activities, two forms were used in this study. Experiment A included an identity-salience oral prompt and results initially proved to be statistically insignificant according to the *t*-test. Experiment B utilized a Japaneseness-survey (i.e., written prompt) for the identity-salience activity portion of the experiment. Results were determined to be statistically insignificant. However, upon breaking down the Experiment A and Experiment B scores by range, one can see a consistent drop in the number of participants that scored in the top of the range in both experimental groups, when compared to the control groups. This fact helps answer RQ 2, and lends support to the primary hypothesis, which aligns with the results of previous studies of stereotype threat.

There were limitations in the present study. Because of educational constraints, the researcher was not able to randomly assign participants to the control or experimental groups. Additionally, the instrument used in the present study was a short, 1-skill (reading) TOEIC quiz. Utilization of a full-length, four-skills test (reading, writing, listening, speaking) would be a better measure of the participants' overall English abilities, and it would likely yield more robust evidence regarding the relationship between stereotype threat and second language learning. Also, as the study was preliminary in nature, there was no control for the participants' investment or motivation in English language learning in the methodology. Future studies should give consideration for this, as it would likely yield more conclusive and accurate results. Considering the preliminary nature of this study, some trends remained consistent with findings in prior studies, and so there is some evidence to support the primary hypothesis. However, further investigation into the role of stereotype threat in L2 acquisition in Japan is warranted.

It is important to address what can be done to mitigate the negative effects of stereotype threat. The inference between stereotype threat and L2 acquisition in the present study was weak, but it deserves further investigation. For this reason, the researcher did not explicitly address the issue of stereotype threat mitigation. It is appropriate first to grasp the nature of the problem before one can truly consider how to mitigate it. It may be important to note that several implications for educators can be found in previous studies of stereotype threat. In general, it can be said that promoting a general awareness of stereotype threat and the barrier that it imposes on certain groups is one step toward mitigating those negative outcomes. Ackerman-Barger et al. (2016) conclude their study by claiming that "supportive relationships between students and faculty can nurture positive learning outcomes" (p. 1242). Faculty members are often in a good position to mentor students when stereotype threat may be a risk-factor. Giving high-quality feedback is imperative for students suffering under stereotype threat. They write that "when students trust the feedback given, they are more likely to try to implement suggested changes" (p. 1243). Tomasetto et al. (2011) likewise concluded by noting that parents should not endorse negative stereotypes, as children are able to subconsciously pick up on their parents' beliefs, which leads to increased and prolonged stereotype susceptibility among the general population. Finally, it may be important here to mention the antithesis of stereotype threat. This is called a stereotype tax. This is "when a stereotype that others have about you works to your advantage" (Vedantham, 2015, para. 17). Further research concerning this concept may also lead towards identifying effective methods of mitigating the negative effects of stereotype threat in SLA and contribute to a fuller understanding of language learner identity.

References

- Ackerman-Barger, K., Valderama-Wallace, C., Latimore, D., & Drake, C. (2016). Stereotype threat susceptibility among minority health professions students. *Journal of Best Practices in Health Professions Diversity*, 9(2), 1232-1246. <https://www.istor.org/stable/26554256>
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effect of identity activation on quantitative performance. *Psychological Science*, 12(5), 385-390. <https://doi.org/10.1111%2F1467-9280.00371>
- Aronson, J., Quinn, D., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). Academic Press.
- Barkhuizen, G. (2018). Research in English for speakers of other languages. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley. <https://doi.org/10.1002/9781118784235.eelt0227>
- Block, D. (2007). *Second language identities*. Continuum.
- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, 39(6), 626-633. [https://doi.org/10.1016/S0022-1031\(03\)00039-8](https://doi.org/10.1016/S0022-1031(03)00039-8)
- Chaffee, K. E., Lou, N. M., & Noels, K. A. (2020). Does stereotype threat affect men in language domains. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.01302>
- Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588-594. <https://doi.org/10.1177%2F0146167298246003>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.
- Educational Testing Service (2018). *Test and score data summary for 2018 TOEFL iBT tests*. https://www.toefl-ibt.jp/dcms_media/other/toefl_data.pdf
- Gardener, R. C., & MacIntyre, P. D. (1993). A student's contribution to second language learning. Part II: Affective variables. *Language Teaching*, 26(1), 1-11. <https://doi.org/10.1017/S0261444800000045>
- Hermans, H., & Hermans-Konopka, A. (2010). *Dialogical self theory*. Cambridge University Press.
- Kayi-Aydar, H. (2018). Reading silence in Japanese classrooms. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley. <https://doi.org/10.1002/9781118784235.eelt0799>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Kutuk, G. (2019). *The effects of stereotype threat on foreign language performance through the mediating roles of self-efficacy and anxiety* [Unpublished doctoral dissertation], Edge Hill University. https://research.edgehill.ac.uk/ws/portalfiles/portal/20610770/Gulsah_Kutuk_PhD_Thesis_with_Corrections.pdf
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th Ed.). Oxford University Press.
- Muzzatti, B., & Agnoli, F. (2007) Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, 43(3), 747-759. <https://psycnet.apa.org/doi/10.1037/0012-1649.43.3.747>
- Norton, B. (2017). Learner investment and language teacher identity. In G. Barkhuizen (Ed.), *Reflections on language teacher identity research*, (pp. 80-86). Routledge.
- Pavlenko, A., & Blackledge, A. (Eds.), (2004). *Negotiation of identities in multilingual contexts*. Multilingual Matters.
- Pavlova, M. A., Weber, S., Simoes, E., & Sokolov, A. N. (2014). Gender stereotype susceptibility. *PLOS ONE*, 9(12), 1-13. <https://doi.org/10.1371/journal.pone.0114802>
- Peirce, B. N. (1995). Social identity, investment, and language learning. *TESOL Quarterly*, 29(1), 9-31. <https://doi.org/10.2307/3587803>
- Picho, K., & Brown, S. W. (2011). Can stereotype threat be measured? A validation of the social identities and attitudes scale (SIAS). *Journal of Advanced Academics*, 22(3), 347-411. <https://doi.org/10.1177%2F1932202X1102200302>
- Rothgerber, H., & Wolsiefer, K. (2014). A naturalistic study of stereotype threat in young female chess players. *Group Processes & Intergroup Relations*, 17(1), 79-90. <https://doi.org/10.1177%2F1368430213490212>
- Sasayama, S. (2018). Learner characteristics, individual learner differences, and learner role. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley. <https://doi.org/10.1002/9781118784235.eelt0214>
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38(2), 194-201. <http://dx.doi.org/10.1006/jesp.2001.1500>
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83. <https://doi.org/10.1111%2F1467-9280.00111>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28. <https://doi.org/10.1006/jesp.1998.1373>
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613-629. <https://psycnet.apa.org/doi/10.1037/0003-066X.52.6.613>
- Steele, C. M. (2010). *Whistling Vivaldi: How stereotypes affect us and what we can do*. Norton.
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811. <https://psycnet.apa.org/doi/10.1037/0022-3514.69.5.797>
- Thurman, J. (2018). Affect. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley. <https://doi.org/10.1002/9781118784235.eelt0201>
- Tomasello, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology*, 47(4), 943-949. <https://psycnet.apa.org/doi/10.1037/a0024047>
- Toohey, K. (2001). *Learning English at school: Identity, social relations and classroom practice*. Multilingual Matters.
- Vedantham, S. (Host). (2015, October 6). An ace up the poker star's sleeve: The surprising upside of stereotypes. *Hidden Brain* [Audio Podcast]. NPR. <https://www.npr.org/2015/10/06/444213368/an-ace-up-the-poker-stars-sleeve-the-surprising-upside-of-stereotypes>