A Meta-Analysis of the Relationship Between Study Abroad and Intercultural Competence

by

Jeffrey David Burrow

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Leadership, Higher, and Adult Education
Ontario Institute for Studies in Education
University of Toronto

A META-ANALYSIS OF THE RELATIONSHIP BETWEEN STUDY ABROAD AND INTERCULTURAL COMPETENCE

Jeffrey David Burrow

Doctor of Philosophy

Department of Leadership, Higher, and Adult Education
Ontario Institute for Studies in Education, University of Toronto

2019

Abstract

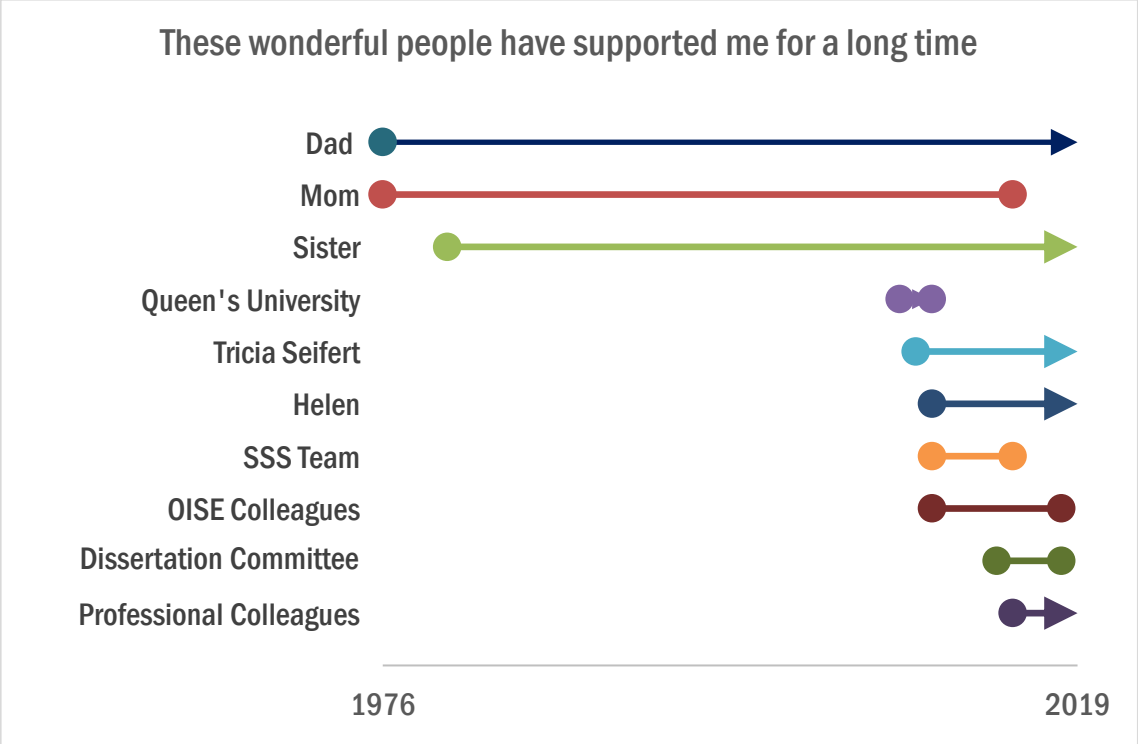
A growing body of research dating back to the early 1980's has focused on the outcomes and impacts of college and university students participating in study abroad programs. A large proportion of this research has examined how intercultural competence is enhanced for students from participating in study abroad. Intercultural competence is valued as an interpersonal skill for personal and civic responsibilities, but it is also believed to enhance employability and effectiveness in the labour market. Despite decades of research, there have been few attempts to quantify the relationship between study abroad and intercultural competence and fewer still to investigate which student, study abroad program and research design characteristics are associated with this relationship.

In this dissertation, studies that included a pre-test and post-test measure of intercultural competence in a study abroad program were reviewed and coded for a series of student, study abroad and research design characteristics aligned with Engle and Engle's study abroad program framework and Astin's Inputs-Environments-Output model. Ultimately, 72 studies with 85 unique results were used. A random-effects meta-analysis approach was used to combine the results of all these studies into an overall effect size of $g = 0.38$, $df(84)$, 95% CI (0.32, 0.45), $p < .001$ which is small to medium effect. Using a multivariate meta-regression approach it was found that studies that

included large proportions of students in business or STEM programs, a foreign language component, had meaningful cultural interaction or used a retrospective pre-tests were associated with larger effect sizes. In contrast, studies with larger proportions of social science students, studies that used the Intercultural Development Inventory or the Global Perspectives Inventory studies were associated with lower effect sizes.

By quantifying the magnitude of the overall relationship and testing individual and sets of moderator variables, this study may provide guidance for those directly involved in planning and organizing study abroad programs, institutional leaders and policy makers who fund study abroad programs and want to increase the number and profile of participants, and for those who research study abroad, intercultural competence and other outcomes of interest.

Acknowledgements



I am not entirely sure I would ever get to this page of my dissertation, but here I am, and there was no way I wouldn't include at least one dataviz in it. This dissertation took a long time and I owe a lot of thanks to a lot of people. First to my dissertation supervisor Dr. Tricia Seifert, thank you for supporting me even before you were a faculty member at OISE. You are a gifted teacher and created an incredible research opportunity for me and many others on the Supporting Student Success team. The amount I've learned from you would fill a dissertation at least as long as this. Thank you for the guidance and mentorship and being great company till the end. I am also very grateful to my committee members Dr. Hayhoe and Dr. Chmielewski for their thoughtful and challenging feedback the entire way through. A huge thanks to everyone I met and learned from at OISE, but especially those on the SSS team with Tricia - Jacquie, Kathleen, Diliana, Kim, and Christine. What an amazing team and experience that was.

I would not have made it to a doctoral program without the incredible classmates I met at Queen's during my masters. And I want to especially thank Dr Don Klinger who supervised my master's thesis and Dr Jennie Massey who gave me my first chance to do research on study abroad.

Since transitioning from full-time student to full-time staff, I've had a wonderful opportunity to work with, and learn daily from, a huge number of incredibly smart, thoughtful and caring student affairs pros who have helped me be better at my job and a better doctoral student. Thank you to David and Adam for bringing me into the most fascinating job I could have possibly imagined. Thank you Kim, Sania, and Atifa (who checked in weekly with me for nearly 3 years!). You are three wickedly smart and inspiring young professionals and I can't wait to read your dissertations😊 Finally, a big thank you to my assessment and higher ed people, Erin, Sara, and Lesley. It's amazing how much I've learned from you in such a short time, even though most of our chats are online and on conference calls.

The final thanks go to my family. My sister Leslie (who gave me a place to sleep when I started this PhD) and whose house, and kids, always provided a wonderful break away from school. Thank you to my dad, Wayne, who started by helping me truck all of my belongings from Kingston for my Master's to start this program and encouragement everyday since. Thank you to my mom, Brenda, who passed away in the fall of 2014, but showed all us what strength, perseverance and humility looks like. The last thank you goes to my partner Helen. We met just as I was finishing my master's. For all the times in the last 9 years I stayed an hour or two or four longer at the library than planned. All the weekends I was gone early in the morning. And the many times, usually late at night, I

told you I wanted to quit. You never said, ‘No! Why? You’re so close’. You simply asked me why I was feeling that way and what I wanted to do. The highs and lows during this time have been more than either of us could imagined. But there is simply no way I would have finished this without your support, encouragement and love. Today is basically the first day I’ve not been a student during our relationship. And it’s kind of exciting to think about what’s next.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Whether for two weeks, a month, a semester or a full academic year, the opportunity to study abroad is a unique opportunity for undergraduate students; one that is often described as formative and life-changing. Around the world, a growing number of students are seeking out opportunities to study abroad. For example, in 2016-2017 an estimated 30,000 Canadian, 325,000 American and 300,000 European students studied abroad within the EU alone (Canadian Bureau of International Education [CBIE], 2016; European Union, 2017; Institute of International Education, 2016). Students interested in studying abroad have a myriad of options from almost any academic program of study in nearly every corner of the world. Though the ability to receive academic credit for courses abroad is an attractive feature - and one that is often necessary to justify the considerable cost - industry leaders, policy makers, lobby groups and post-secondary institutions actually more often argue for greater participation in study abroad based on the numerous ways in which students develop and grow during their time abroad (Advisory Panel on Canada's International Education Strategy [APCIES], 2012; Association of American Colleges and Universities [AACU], 2007; Hoffa & DePaul, 2010; Twombly, Salisbury, Tumanut, & Klute, 2012). Research findings have linked study abroad participation with a broad range of areas in student development, including foreign and additional language development (Varela & Gatlin-Watts, 2014; Yang, 2016), personal development (Forsey, Broomhall & Davis, 2012; Gmelch, 1997; Milstein, 2005.), morals and values (Pendleton, Cochran, Kapadia & Iyer, 2015; Ryan & Twibell, 2000), identity development (Craig, Zou, & Curtis, 2018; Dolby, 2004/2007; Kinginger, 2013; Pitts, 2009), cognitive or intellectual development (Cai, & Sankaran,

2015; McKeowen, 2009; Sung, 2016) and intercultural competence (Chieffo & Griffiths, 2009; Clarke, Flaherty, Wright, & McMillen, 2009; Kitsantas, 2004; Varela, 2017).

Of these findings, it is the research relating to intercultural competence that is most commonly advanced in Canada and the United States as an argument for why students should study abroad, and institutions devote considerable space to advertising the study abroad opportunities available to students at their campuses (Lukosius & Festervand, 2013; Woolf, 2007; Zemach-Bersin, 2007/2009/2012). Study abroad offices are adorned with posters, brochures and other marketing materials which promote the dozens, if not hundreds of study abroad partners and programs their students can take part in. Senior institutional leaders and their lobby groups reiterate the importance of intercultural experiences, echoing calls from industry who underscore how much they value graduates who can interact with individuals from different cultural backgrounds. These attempts by various stakeholders to advocate for more participation are underpinned by dozens and dozens of research studies that have attempted to highlight that 'going abroad' leads to intercultural development (see Ogden & Streitwieser, 2016; Salisbury, 2011; Twombly et al., 2012; Varela; 2017 for summaries). Journal articles, dissertations, technical reports, and conference presentations, with few exceptions, conclude with statements that study abroad is an effective way for students to develop intercultural competence.

Despite these claims, what remains unanswered is the magnitude of this intercultural competence development, and a nuanced look at whether, and to what degree this development may be unique to certain populations of students, characteristics of study abroad programs, or survey instruments and research designs. Answering those

questions does not necessarily require another study of students abroad and students who study at home. Instead, this dissertation aims to utilize the existing research of those who have researched study abroad and intercultural competence to understand both the magnitude of the effect and the student and study abroad program and research design characteristics that have the greatest influence on intercultural competence.

Definitions

There are two primary constructs of interest in the proposed study; study abroad and intercultural competence. Before continuing further in this dissertation, it is pertinent to define them both.

Study abroad. Study abroad is used to describe a wide range of activities that occur internationally. It can describe many credit-earning experiences that require international travel activities like internships, teaching placements, research, international service learning or field studies (see Bond et al., 2009). While it is likely that these activities have significant and numerous benefits, the focus of this study is on classroom-based study abroad programs. In this dissertation, study abroad is defined as "educational programs of study, in countries outside the student's home institution, that result in progress towards an academic degree at a student's home institution," (Forum on Education Abroad, 2018).

Intercultural competence. Researchers have commented on the dozens of definitions and models of intercultural competence (Deardorff, 2006/2011; Leung, Ang, & Tan, 2014; Sinicrope, Norris, & Watanabe, 2007; Sercu, 2010; Spitzberg & Changnon, 2009). Several of these definitions and models will be examined in the literature review chapter. For the purposes of this study, intercultural competence (ICC) is defined as "the

ability to relate and communicate effectively when individuals involved in the interaction do not share the same culture, ethnicity, language, or other common experiences" (Forum on Education Abroad, para 1, 2018).

Globalization, Internationalization and Study Abroad

Contemporary study abroad is the act of a single student studying away from the country and institution where they are enrolled for their degree. However, it is just one element of a much larger systematic institutional and national response to globalization (Knight, 2004, 2012, 2014). For governments at a federal, state or provincial level, study abroad programs are believed to promote the brand and reputation of their education systems globally and enhance economic competitiveness considering increasing globalization pressures (APCIES, 2012; Bond et al., 2009; Foreign Affairs, Trade and Development Canada, 2014; Hudzik, 2015; MAESD; 2018; Australia National Strategy for International Education 2025, 2017; U.S Department of Education, 2014). At the institutional level, study abroad is viewed as a key element of their internationalization strategy, aiding them in globalizing the student experience, deepening their international partnerships, and improving their standing in international rankings (AUCC, 2007; Altbach, 2012; Hudzik, 2011, 2015; Taskoh, 2014; Nolan & Hunter, 2012; Sutton, Egginton, & Favela, 2012).

For students, study abroad is viewed as an opportunity to develop intercultural competence, become more independent and take courses in a country and context not available at home; all of which may help position them better for professional and civic responsibilities upon graduation (Massey & Burrow, 2012; Paige, Fry, Stallman, Josić, & Jon, 2009; Pope, Sánchez, Lehnert & Schmid, 2014; Tillman, 2012). Given the variety of

stakeholders who benefit from study abroad programs, it is not surprising that there have been growing calls for more students to participate.

Study abroad was initially conceived as an opportunity for a very elite and very small number of students to gain linguistic and academic experience while being enrolled 'overseas' (Hullihen, 1928). While still only a small percentage of student participate study abroad has become an exemplar of the growing role that globalization and internationalization have in modern post-secondary education systems. In thinking about how study abroad 'fits' within the larger post-secondary education enterprise, it is important to consider two additional forces that are shaping not only educational, but broader societal movements as well: globalization and internationalization.

Globalization. Though definitions are contested, in this paper globalization is considered "the economic, political, and societal forces pushing 21st century higher education toward greater international involvement," (Altbach & Knight, 2007, p. 290). In using this definition, the impacts of globalization are evident in many aspects of higher education. These range from the commercialization of academic research, the growing use of English in both research and teaching, the movement of scholars and students across borders, the commodification of higher education and the growing focus on international rankings and pressures on systems to develop and institutions to become 'world-class' (Altbach & Knight, 2007; Deardorff, Wit, Heyl & Adams, 2012; Hazelkorn, 2015; Marginson & van der Wende, 2009; Salmi, 2009). Globalization is embodied in the external, larger societal forces and pressures exerted on nation states, institutions and individual faculty and students.

Internationalization. In contrast, internationalization involves the actions and choices made by these very same stakeholders in higher education considering, and in response to, these globalizing forces and trends. It is the "process of integrating an international, intercultural or global dimension into the purpose, functions or delivery of post-secondary education," (Knight, 2004, p. 11). Previously, scholars characterized the drive behind internationalization by academic (international dimensions to research and teaching), economic (revenue generation from international student tuition or economic competitiveness), political (foreign policy and national security) and social (intercultural understanding) rationales (Knight, 2012). Reviewing these rationales and the examples of activities for each highlight that there are national (or system-level) actions alongside institutional activities. Thus, a more robust model would acknowledge the role of both system-level and institution-specific responses and international activities. Defining internationalization in this way positions it as a response of governments, post-secondary institutions, and students to these current or future changes arising from globalization.

***Internationalization of colleges and universities*.** At the institution level, strategic planning around internationalization is increasingly common. Research indicates a growing trend towards internationalization plans and creating senior leadership positions within their colleges or universities dedicated to 'international' on their campus (de Wit, Hunter & Coelen, 2015; Hudzik & Stohl, 2012; International Association of Universities, 2010, Knight, 2006; Seeber, Cattaneo, Huisman, & Paleari, 2016).

At an institutional level, internationalization activities include both those that occur 'at-home' and abroad (Knight, 2012). Internationalization at home can include

many types of activities. One aspect includes the development of more 'internationally-minded' curriculum for students. The intent is for the curriculum at home to take on a more international flavour and be more aware of the world-wide scholarship that exists and the global and broader societal impacts on each academic discipline. The OECD characterizes this as "curriculum with an international orientation in content and/or form, aimed at preparing students for performing (professionally/socially) in an international and multicultural context and designed for domestic and/or foreign students," (1996, p. 9). The potential in these at-home initiatives is significant as curricular modifications and programming can impact hundred or thousands of students at an institution (Brewer & Cunningham, 2010; Brewer & Leask, 2012; Harrison, 2015). Examples could include the development of programs with an international theme or programs focused on area or language studies. It can also be enacted at the individual course level with the introduction of more international content and reflection in a course syllabus. Internationalization at home can also focus on domestic experiences, away from campus. These could involve a group of students travelling with a faculty member to another institution or a different part of the country, to learn in a cultural setting and with people who do not have the same cultural background as typical study abroad participants (Sobania & Braskamp, 2009). These domestic opportunities could still provide a deep cultural experience but might encourage broader profiles of applicants for students without previous international experience, for whom travelling internationally is prohibitively expensive and due to passport and visa status may not be able to leave the country they are studying in (Sobania & Braskamp, 2009). Thus there are opportunities

for increasing the number of students who can have an intercultural experience by broadening the focus of where intercultural experiences happen.

A second element of an institutional internationalization plan focuses on recruitment of international students for full-time study on campus. These recruitment efforts can help to diversify the student body and create opportunities for interaction with more diverse perspectives while providing additional revenue streams for post-secondary institutions and systems (APCIES, 2012; Australia, 2017; Foreign Affairs, Trade, Development; 2014; MAESD, 2018). International student enrolment is increasingly used as a marker for how 'international' a university is and often is an indicator in international ranking systems, such as the Times Higher Education and Shanghai Jiao Tong University rankings (Green, Marmolejo & Egron-Polak, 2012; Hazelkorn, 2015; Ordorika & Lloyd, 2015). More recently, community colleges have entered the international recruitment arena and increasingly numbers of international students have enrolled in their diploma and certificate programs (APCIES, 2012; Amani & Kim, 2017; MAESD, 2018). Regardless of institution type, international students could play a vital role in helping to internationalize curriculum by introducing students and faculty to research, scholars and practices that exist outside of their host institution and post-secondary context, provided that institutions recognize their potential contributions beyond additional tuition dollars.

Finally, we see education abroad programs and opportunities for students as part of internationalization strategies. These can include international service-learning courses and projects, co-operative education abroad, international internships or placements for professional programs (teacher education, engineering and social work for

example) and of course, study abroad (Knight, 2004, 2012; Teichler, 2017). These programs may, broadly considered be called international ‘study away’ programs in that they involve travel, to another country and institution for a period of time to advance their educational experience (Sobania & Braskamp, 2009). The most commonly known type of study away program, study abroad is an activity institutions envision as not just a potential part of a student’s undergraduate degree or diploma experience, but as a key element of both their internationalization efforts and their undergraduate recruitment strategy (Knight, 2012; Massey & Burrow, 2012; University of Toronto, 2017). Though it may only impact a small percentage of undergraduate students at most institutions, study abroad can be a major element of institutional marketing efforts (Zemach-Bersin, 2007, 2009, 2012). Institutions and the staff who represent them are usually quite able to describe how many of their students 'go abroad', what percentage of their students have an international experience and how many agreements, partnerships or consortia arrangements in which their institutions are involved (Engle & Engle, 2003).

The section above highlights three of the most common elements of campus internationalization strategies. They are presented in order of the number of students that are likely impacted or involved by them. Internationalizing the curriculum, either through specific programs, degree options or even specific courses, are internationalizing activities that can impact virtually all students studying at a college or university. The second focus is growing international student enrolment, which is a priority at the system and institutional level (Adams, Leventhal & Connelly, 2012; APCIES, 2012; Australian Government, 2017; Banks & Bhandari, 2012; Hawthorne, 2012; MAESD, 2018). Though the contribution of international student recruitment to internationalization is often

expressed in economic terms, there is a growing focus on what international students can bring to the classroom and the community they live in while studying for their degrees (Belkhodja & Esses, 2014; Lee, 2015; Valedez, 2015). The final strategy is the promotion of education abroad and specifically, study abroad opportunities. Despite the prominent role study abroad plays in undergraduate recruitment efforts, the activity impacts only a small percentage of students at most institutions. Chapters three and five will provide more information on the characteristics of those students (at least in the United States) and of the study abroad programs they participate in. The next section will examine some of the rationales used for study abroad as an internationalization strategy.

Stakeholder Arguments for Promoting Study Abroad Participation

The growing presence of internationalization plans, and their focus on study abroad as a key pillar of government and institutional efforts to promote study abroad are difficult to ignore. However, behind the belief that more students should be studying abroad are a broader set of rationales, arguments and support for why more is needed and why more is better. Recall that study abroad is just one, typically very small, element of internationalization activities. This section investigates the arguments advanced by three sets of stakeholders: various levels of government (national, provincial, or supra-national), post-secondary institutions and finally lobby and advocacy groups.

Government perspectives on study abroad. In Canada, the complex jurisdictional distinctions between federal and provincial levels of government have presented a great challenge in international education for decades (Trilokekar & Schubert, 2009). While education is the responsibility of provincial governments, foreign affairs, international development, and human resources are federal responsibilities

(Jones, 2004). In the early 2000's it was suggested that "no national agenda for international higher education exists. Nor, it seems, is this weakness likely to be remedied soon" (Barrow, Didou-Aupetit, & Mallea, 2003, p. 121). However, in 2012, the Advisory Panel on Canada's International Education Strategy (APCIES) released a report, *International Education: A Key Driver of Canada's Future Prosperity*, through the Department of Foreign Affairs and International Trade and in 2014, *Canada's International Education Strategy* was released (DFAIT, 2012; 2014). The 2012 report focuses primarily on internationalization efforts like international student recruitment and communicating national education quality internationally (APCIES, 2012).

From a study abroad perspective, the most noteworthy element of the report is a goal of having 50,000 Canadians students go abroad by 2022 (APCIES, 2012). This would represent an increase of approximately 60% over 2012 levels at that time (CUSC, 2012). The rationales the APCIES advance focus on cultural learning and skill development. The report acknowledges the importance of ensuring that "students have the intercultural and language skills to become leaders in the global knowledge economy—that they become "global citizens" (APCIES, 2012, p. 36). The report also notes that studying abroad enables students to "learn from diverse perspectives, immerse themselves in diverse cultural practices, and learn foreign languages required to conduct business globally" (APCIES, 2012, p. 36). Although both statements highlight the value of developing an awareness of cultural differences, the Panel suggests this development and learning while abroad is support of economic and career development.

The APCIES report was followed in 2014 by a Foreign Affairs, Development and Trade report. While the APCIES report had a small, but not unimportant, number of

references and goals for the government to focus on as it related to study abroad, the official strategy contained very few references to study abroad and vague notions about increasing the number of students who 'study abroad'. The 2014 report spoke of a desire to "increase the number, breadth and depth of active collaborations between Canadian and foreign post-secondary institutions and research centres, and to position Canada as a country of choice for both academic recruitment and partnerships" (DFAIT, 2014, p. 11). Missing were the ambitious participation targets and in place were statements of that focus mainly on enhancing research and institutional collaborations. There was also a significant emphasis on international student recruitment, national branding, and the importance of creating revenue opportunities through training and development program. In fact, the phrase "study abroad" appeared just twice and one of those is in reference to the 2012 report.

In March 2019, the Federal budget announced a plan to invest nearly CDN $150 million towards international education from 2019-2020 until 2023-24 (Government of Canada, 2019). Some portion of the funds will be dedicated to promoting the education brand of Canada and export promotion; however a large proportion appears to be designated to support international education opportunities. The budget stated that the funding was to support 'outbound student mobility' to support students in Canada who wish to gain international work/study opportunities. While higher education institutions and lobby groups have long sought investments in international education, at least three very important questions remain about the funding and next steps. The first is whether the funds are targeted at working abroad, studying abroad or both. Both have the opportunity to be valuable experiences for students but how this funding is targeted speaks a lot to

what the federal government sees as a priority. Secondly, given that the participants in study abroad in the U.S. are not reflective of the general population of students in higher education either ethnoculturally or socio-economically, I am curious to see if some portion of the funds are targeted towards populations of students who have been historically underrepresented and underserved by study abroad. Similarly, it is not clear if the government will target specific program characteristics. Specifically, is there a minimum length, preference for work/study in an additional language, or targeted countries or regions? Finally, the mechanisms and the ways in which the funding will flow to students or to institutions remain unknown. Will the funds be distributed to all post-secondary institutions (numbering in the hundreds), or might students apply directly to the government or a third party (like CBIE, Universities Canada or Colleges and Institutes Canada)? Undoubtedly, institutions and lobby groups will be watching these developments closely.

In the United States, far more ambitious study abroad plans have been consistently proposed, often via the Senator Paul Simon Study Abroad Act (most recently in 2017). Two previous versions of the bill have passed previously, making available considerable resources for students to study abroad, specifically students from community colleges and minority-serving institutions, and in its current iteration stating a goal of "not less than 1,000,000 undergraduate students studying abroad annually" (Simon Study Abroad Act, S. 601, H.R. 4379, 2017). The document highlights an official bill in Congress that although not currently funded, provides important statements about the critical importance of study abroad participation to the United States (NAFSA, 2017) and sets new and ambitious goals for the number of students that should study abroad. A

related initiative was developed to help increase the number of U.S. students studying abroad in China to 100,000 annually (US-China Strong Foundation, 2018). This initiative is unique in that it has the support of the Chinese government, who is providing financial assistance for 20,000 of these students. Whereas the Simon Study Abroad Act remains a broadly supported but still unfunded initiative, the 100,000 to China program has significant external support from current and former government officials, foundations and business leaders: "China is the second-largest economy in the world, a major strategic power and the United States' fastest-growing trade partner. It is perhaps the world's most consequential bilateral relationship, as virtually every future global challenge will require our two countries to work together. We must invest in this critical partnership." (US-China Strong Foundation, 2018) Thus, at the U.S. federal level, the interest in promoting study abroad appear to be related to an understanding of the benefits to students of living and learning in another academic environment and cultural context from their primary country of study, though the 100,000 to China initiative has a clear focus on national security and economic rationales.

Colleges and universities. A 2016 survey of Canadian universities found that the primary internationalization goals were to enhance international student recruitment, increase the number of students engaged in study away and study abroad programs and strengthen internationalization at home initiatives (CBIE, 2016). At the individual institution level, the level of support for and prominence of study abroad varies considerably. On one hand, institutional marketing materials targeted at high school students often contain references to study abroad. These are usually in terms of the number of exchange partners they have, some of the desirable locations where students

have studied abroad, and pictures of large groups of students in front of clearly recognizable landmarks like the Eiffel Tower, the Great Wall of China or Machu Picchu (Zemach-Bersin 2007, 2009). The CBIE report provides the best overview of how Canadian post-secondary institutions and students in general view study abroad. The report focuses on the low participation rates, estimated to be 2-3% annually in Canada, discusses the challenges in data collection as it relates to the 'counting' of study abroad participation, but provides no real suggestions as to what the priorities are or should be for Canada. It does not seem to recognize the previous work done by the APCIES and the ambitious study abroad targets and calls to action proposed.

Reviewing strategic plans can provide some insight into how institutions and their leaders position study abroad. The University of British Columbia has a specific internationalization strategy as part of its overall strategic plan. Embedded within it is the following:

> Increase student participation in mobility programs so that 30% of all undergraduates at both campuses have an international experience by the time they graduate, and establish a mobility participation goal specific to graduate students (University of British Columbia [UBC] International Strategic plan, 2011, p.11)

UBC sets an ambitious target that represents a doubling of their existing participation. Within their international websites, on the homepage dedicated to global opportunities, they note:

> Introducing yourself to the culture of another region by living, studying and working abroad builds independence, initiative and adaptability - important traits that employers are constantly on the lookout for. (UBC Student Services, 2018)

The first part of the statement is about adaptability and development, but the sentiment turns towards employability as the primary rationale for studying abroad. Another large Canadian university, McGill, in their international strategic framework notes a desire to:

> Increase to 15% the number of undergraduate students undertaking outward mobility by reducing administrative barriers to credit transfer and expanding credit-bearing opportunities (McGill University International Strategic Framework).

Here McGill also sets a specific target, but there are few details about how they intend t move from the existing state to the bigger number. However, McGill discusses with their international framework a:

> potentially invaluable opportunity to strengthen McGill's international and national reputation as a university whose graduates are seen as model global citizens in an increasingly plural world, based on principles of mutual respect, open-mindedness, and an appreciation of alternative understandings of the world that are consistent with fundamental conceptions of human rights (McGill University International Strategic Framework, 2019).

McGill takes a very broad view of the value and purpose of internationalization and discusses graduates as global citizens, as opposed to those with greater employability.

Western University (London, Ontario) includes internationalization as one of its primary principles in their strategic plan and says that "We will embrace our role as an active member of the global academic community through the full range of our educational, research, scholarship, and community development activities that engage our students, faculty, staff, alumni and external partners" (Western University, 2018) With respect to study abroad and international opportunities, Western denotes a desire to "Enhance the learning experience by providing a community-based experiential learning opportunity, an international learning opportunity, or a research learning opportunity for all undergraduates who wish to pursue one as part of their degree" (Western University, 2018).

Unlike UBC and McGill, Western does not commit to a specific target for study abroad participation, instead denoting it as something students can participate in as a form of experiential or global learning if they wish. Western also does not advance why

they are supporting the development of international opportunities and study abroad.

Finally, Queen's University (Kingston, Ontario) has internationalization embedded within its strategic plan, rather than being a separate section. In their most recent strategy document, there is a section that focuses specifically on exchanges and other international education opportunities. Their plan states:

> If international partnerships are to be meaningful and students to be prepared for international experience, they must be informed by an educated appreciation and knowledge of other languages, histories, cultures and their contributions to a shared modernity and humanity that allows our students to learn from others as well as to teach or assist (Queen's University, 2011).

These are four of the largest and highest ranked institutions in Canada. (Note that the University of Toronto does not currently have a defined internationalization strategy). These represent a range of perspectives on the value of study abroad and internationalization. How these statements are translated into strategy, tactics and financial support for staff and faculty involved in, or who wish to develop study abroad opportunities is not known.

Lobby and advocacy groups. Unlike the two previous stakeholders, lobby and advocacy groups differ by existing in spaces between the institutions and the government. They are often funded by colleges, universities and corporations to influence public opinion, advance the interests of their membership and act as an advocate to various levels of government. In Canada, the Canadian Bureau of International Education (CBIE) is the most notable lobby group for international education in Canada. On their website, CBIE states that their purpose is to be “the national voice advancing Canadian international education by creating and mobilizing expertise, knowledge, opportunity and

leadership. CBIE is a global leader in international education, dedicated to equity, quality, inclusiveness and partnership" (CBIE, 2018).

The CBIE mission and vision statements do not include the words 'abroad' or 'students'. Neither do their value statements nor statement of operational principles. Only a vague reference to that "CBIE is a global leader in international education, dedicated to equity, quality, inclusiveness and partnership," (CBIE, 2018). Though not explicitly in their guiding documents, CBIE appears to be currently involved in coordinating a Canadian Learning Abroad network, focussed on responding to some of the key issues raised by students in the institutional survey from 2016 (CBIE, 2018). This includes attempting to lobby and argue for more scholarships and bursaries to defray participation costs, coordinate promotion across Canada for study abroad, and help normalize the idea of 'going abroad'. What is less evident from their website is why CBIE supports study abroad, and why they feel institutions should develop more opportunities and more student should study abroad.

In contrast, NAFSA: The Association of International Education, an American equivalent of CBIE, notes that its mission is to focus on four key elements of institutions' internalization efforts. These are to advance public policy for international education, internationalize higher education, support international students and scholars and finally to support study abroad. Their website focuses on all aspects of the study abroad experience from program development to measuring outcomes, ensuring safety, orientation and reintegration (NAFSA, 2018). The differences between the two groups are noteworthy. While CBIE describes its role as promotion of the Canadian education brand and the recruitment of international students, NAFSA clearly embraces a mission

that encompasses many elements of internationalization. CBIE supports study abroad by including sessions in its annual conference (as does NAFSA) but does not appear to aid member institutions in developing and enhancing their study abroad efforts. NAFSA has taken a clear leadership role in sharing best practices, compiling and developing resources, policy arguments and position papers related to study abroad.

In addition to NAFSA, several other groups in the U.S operate as advocates for international education and study abroad. The Institute of International Education (IIE), the same organization that publishes the annual report on U.S. students abroad and international students in the US, are leading efforts to help encourage 100,000 American students to study abroad in China by 2022. IIE notes that "Globalization is changing the way the world works, and employers are increasingly looking for workers who have cross-cultural competence and cutting-edge technical skills" (IIE, 2018). IIE highlights the importance of cross-cultural competence as a means for students to be more employable in the workforce. Other rationales they suggest include developing global skills, which can open personal and professional opportunities. Again, IIE positions study abroad from both economic and personal development perspectives.

The Forum on Education Abroad (FEA) is likely the premier group focused on research and assessment of study abroad in the United States. They host an annual conference, publish a peer-reviewed journal and emphasize sharing assessment resources among study abroad programs. FEA states that it "develops and disseminates comprehensive standards of good practice, resources and training, advocates for education abroad and its value, and engages the field in critical dialogue to benefit students," (FEA, para 1, 2018). There is a clear focus in this statement on students, and

helping institutions develop high quality study abroad programs for students to participate. They go on to state:

> Through education abroad, students of every background benefit from their experiences with other peoples, cultures, and countries, in ways not available on their home campus. Together these experiences advance students' personal, academic, and professional growth. In return, host communities and home institutions benefit from students participating in education abroad (FEA, 2018).

FEA, unlike all other groups here, make specific references to an encouragement for all students to participate because of the opportunities to learn from difference in ways that support their personal, academic and career development. They also denote the potential reciprocal benefits of participating in study abroad for students at home.

Summarizing rationales for study abroad. Stakeholders provide a broad range of rationales and arguments for why study abroad programs exist, why they promote them to their students and why they feel more students should participate in them. In their review of the history of American study abroad since the 1960s, Hoffa and DePaul (2010) posit that for the preceding 50 years, the reasons for participating in study abroad have centered on curricular, cross-cultural, career and developmental arguments. The curricular argument focuses on how study abroad "enriches and diversifies undergraduate education, by offering academic learning of a sort not possible on the whole campus, yet of a standard worthy of home campus academic credit," (p. 8). The cross-cultural argument looks at how study abroad can be an opportunity to learn about, live and study in another culture, and at the same time learn about their own 'Americanness' (p. 9). The third is centered on career enhancement in that students' time abroad better position them in the eyes of employers upon graduation. The final argument is a developmental one. This is related to material presented earlier in this chapter that indicated that study abroad

can help students learn and develop in broader intellectual ways, enhance maturity, independence and increase their self-confidence (Hoffa & DePaul, 2010). Though stakeholders advance many reasons to promote study abroad, empirically, the bulk of the published research has fallen into two areas; language acquisition and intercultural competence. The latter is the focus of this study. For more information on language acquisition see (Isabelli-García, Bown, Plews & Dewey, 2018; Segalowitz & Freed, 2004; Yang, 2016).

The State of Intercultural Competence and Study Abroad Research

The previous section summarizes some of the major rationales underpinning government, institutional and lobby group support for study abroad participation. Throughout their arguments are references about the benefits, typically how the development of intercultural competence may aid students professionally and personally. A cursory review of the research both in the conclusions of individual studies as well as literature reviews suggests that study abroad is a very effective mechanism for developing intercultural competence among post-secondary students (Black & Duhon, 2006; Braskamp, Braskamp, & Merrill, 2009; Chieffo & Griffiths, 2004; Clarke, Flaherty, Wright, 2009; Fry et al., 2009; Ingraham & Peterson, 2004; Kitsantas, 2004; Metcalfe, 2007; Sutton & Rubin, 2010; Vande Berg et al., 2009; Zarges, 2016). With few exceptions (see Burrow, 2010; Keefe, 2008; Sell, 1983) research into the relationship between study abroad and intercultural development report positive associations. However, within just the handful of reports noted above are studies that have populations with less than 50% female students (Clarke et al., 2009) and studies with more than 80% female students (Anderson et al., 2006; Williams, 2005). Some studies focus on students

from a specific program of study like commerce (Anderson et al., 2006; Black & Duhon, 2006) or communications (Williams, 2005), while others include students from a wide variety of academic disciplines (Braskamp et al., Salisbury, 2011; Luchesi, 2014). These studies also vary in other student characteristics like previous experience abroad and year of study.

A review of the literature also highlights that the study abroad programs vary in many ways. There are study abroad programs that are just a few weeks long in the summer or during breaks between semesters (Kitsantas, 2004; Luchesi, 2014; Patterson, 2006), studies that focus on programs that are a semester or full academic year in length (Burrow, 2010; Clarke et al., 2009; Pachmeyer, 2014) and studies that contain a variety of program lengths (Ingraham & Peterson, 2004; Salisbury, 2011; Sutton & Rubin, 2010). Study abroad programs can also vary in the type of housing provided (homestay or residence), the academic context of local or home country faculty delivering the courses, the degree of interaction with locals or planned intercultural activities; all factors which could influence the relationship between study abroad and intercultural competence. Many of these studies cite, either in their literature review or their conclusions, an article by paper by Engle and Engle who argued that it was "time to draw distinctions of a qualitative sort—time for international education professionals to consider seriously the elaboration and adoption of one such guide, a hierarchical classification of program type," (2003, p. 2). Engle and Engle outlined a way of thinking about the aspects that differentiate study abroad programs from each other in a range of ways including program length, language of instruction, housing, and cultural interaction and encouraged

researchers to incorporate these differences into their research design and analytical approaches.

Outside of student and study abroad program characteristics, the published research varies in their research design. For example, there are studies that use standardized survey instruments like the Intercultural Development Inventory (Anderson et al., 2006; Nam, 2006; Reza, 2015), the Global Perspectives Inventory (Braskamp, Braskamp, & Merrill, 2009; Burrow, 2010; Doyle, 2009; Gaia, 2015), or the Cross-Cultural Adaptability Inventory (Black & Duhon, 2006; Edwards, 2009; Kitsantas, 2001; 2004; Zielinski 2007) while others have developed their own survey instruments (see Chieffo & Griffiths, 2004; Ingraham & Peterson, 2004; Zarges, 2016). These instruments have been developed using varying approaches and degrees of rigour in terms of construct and content validity. In terms of research design, some use only a post-test design (Hadis, 2005; Ingraham & Peterson, 2004; Zielinksi, 2007) while others utilized a pre-post design (Burrow, 2010; Golay, 2006, Kehl, 2009; Smart, 2014), some a pre-post design with a comparison group (Salisbury, 2011; Sutton & Rubin, 2010; Vande Berg et al., 2009; Zhai, 2000) and a few a retrospective pre-test, post-test design (Armfield, 2004; Palmer, 2009).

Varela Meta-Analysis. During the preparation of this dissertation, a meta-analysis was published using very similar research questions. Varela (2017) attempted to summarize not just the intercultural competence studies (n = 30) but the language acquisition (n = 33) and behavioural outcomes too (n = 11). The researcher did not focus on any student level characteristics, instead using moderators of study abroad program design including program duration, nature of the study abroad program and the housing

format and research design (repeated measures and between groups) (Varela, 2017). The study focuses on some aspects of the Engle and Engle framework, but does not focus on guided reflection and cultural interaction. Moreover, the Varela study includes numerous studies that used a single measurement of two independent groups, which, from a design perspective invites numerous threats to internal validity, specifically selection effects. It is quite possible that these two groups differed meaningfully before the study abroad program began. Thus without pre-test and post-test measures, the author is assuming the groups were similar before the study abroad program and that the only difference that could be related to a post-test difference in intercultural competence is the study abroad program itself. The inclusion of these studies calls into question the validity of the overall estimates.

Study Abroad Research. As noted earlier, most research concludes that there is a positive relationship between study abroad and intercultural development. However, there is also considerable diversity in the student populations studied, the study abroad programs under analysis and the research designs utilized in these studies. Few studies consider how student characteristics, study abroad program characteristics and research design choices are related to their findings. Thus, it is not known to what degree the relationship between study abroad and intercultural development is effectively the same for all participants, regardless of their own characteristics, the study abroad program they participated in or the way the research study was conducted.

Study abroad research tends to focus on whether the relationship between studying abroad and intercultural development was significant. They focus on whether the *p*-value is above or below 0.05. The *p*-value is a result of a statistical test that

estimates the probably of obtaining a result as large or larger than the one observed, if the null hypothesis were true (Gelman & Weakliem, 2009). Social science researchers tend to use a *p*-value of 0.05 or smaller to indicate that an estimate or a difference has reached statistical significance. The history and rationale for why .05 has become commonly accepted may be traced to Fisher who stated that

> the value [*t*-statistic] for which $p = 0.05$, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant (1925, p. 46).

Returning to the studies in this dissertation, if researchers conducted a study in which students were surveyed before and after the study abroad program, they would use the *p*-value as the indicator of if there was a significant change in students' self-reported intercultural competence abilities during their time abroad. The interpretation of this finding would be that it is quite unlikely that a difference of that size was due to chance and that students' intercultural competence did improve while abroad. The *p*-value only tells the researcher and readers that there is a difference; it does not provide any direct evidence of how large or meaningful the difference was. Effect sizes quantify differences using a standardized measure that can be compared (Coe, 2002). Study abroad research is not unique in often misconstruing *significant* as a synonym for important or meaningful (Gelman & Weakliem, 2009). Relying on the term 'significant', ignores the vastly more important issue of the magnitude of the relationships and how student, program or research design characteristics may be associated with the magnitude of the effect.

Despite decades of research, and hundreds of studies that have examined study abroad and intercultural development, very basic notions of what student, program and research design elements influence and impact the magnitude of this relationship remain

unclear. Given the volume of research studies in journal articles, dissertations, technical reports and conference presentations on these two constructs, a review of this existing research may provide valuable practical and research insights into what influences the relationship.

Research Purpose & Research Questions

The purpose of this study is to examine the relationship between participation in study abroad and intercultural development. This will include a comprehensive review of empirical articles, doctoral dissertations, conference presentations, and technical reports that measure study abroad outcomes. These data will be used to answer three research questions:

1. Does an overall relationship exist between study abroad participation and intercultural competence? Is there significant variation in this relationship across studies?

2. To what extent do student population characteristics moderate the relationship between study abroad and intercultural competence? To what extent do study abroad program characteristics moderate the relationship between study abroad and intercultural competence? To what extent do research design characteristics moderate the relationship between study abroad and intercultural competence?

3. To what degree does the student, study abroad program or research design characteristics explain variance in the mean study effect size?

Research Significance

This study will have practical and research implications for study abroad program advisers, study abroad program developers and organizers, as well as the broader research community. Though considerable discussion has focused on what students might be best positioned to study abroad I argue that this should not be a primary focus of those

involved in study abroad programs. It is akin to asking who is best positioned to benefit from post-secondary education.

For those who advise and select students for study abroad programs this research may help to provide a greater understanding of the relationship between student characteristics and intercultural development. This study may identify the degree to which student characteristics are associated with intercultural development. This is relevant as it could be valuable in helping advisers determine which factors to consider in their application and eligibility process.

This study may also provide practical implications for those who develop and manage study abroad programs at operational or strategic levels by identifying how different types and features of study abroad programs affect the relationship between study abroad participation and intercultural development. Characteristics of a study abroad program (e.g. length, housing arrangements, curricular focus and intercultural activities or mentoring) may support or inhibit student intercultural development.

Finally, this study may provide two sets of insights for the research community. The first is an understanding of how the magnitude of effects on intercultural development varies depending on which survey instrument is used. This may provide an ability to compare how sensitive the results are to different survey instruments. Secondly, it will help understand how research designs impact the relationship between study abroad and intercultural development or the extent to which studies that use pre-post or pre-post and comparison groups differ in estimates of intercultural development.

Notwithstanding these potential contributions, this study approaches study abroad, reviews and codes study abroad literature, and considers implications from a primarily

North American perspective. Though every attempt was made to identify and review literature from a global perspective, given the nature of publishing and access to research, whether in journals, dissertations, conferences or books, and my inability to read research in a language other than English, the conclusions are limited to the same geographic regions and systems from which the research was sourced.

Chapter 2: Theoretical and Guiding Frameworks

The proposed study utilizes four theoretical and guiding frameworks. The first is Allport's Theory of Intergroup Contact (1954), which posits that contact with individuals from different groups may lead to reduced prejudice and an improvement in inter-group relations. The second framework in this study is a broad overarching theory of validity, including construct validity, internal validity and external validity (Kane, 2001; 2006, Messick, 1989). The third framework is Astin's Inputs-Environments-Outputs (IEO) model (Astin & Antonio, 2012). This model provides a framework for both considering and later analyzing the relevant student characteristics (Inputs) and study abroad program characteristics (Environments) that can impact intercultural competence development (Outputs). The final piece builds off the IEO model, specifically the environmental (or 'E') by utilizing a theoretical piece by Engle and Engle who detailed seven ways in which study abroad programs could vary substantively from each other (2003). These four frameworks will guide this study.

Allport's Theory of Intergroup Contact

The notion that by studying abroad or 'going abroad' students will develop greater intercultural competence is part of what has been termed the dominant narrative of study abroad. Allusions and direct references to this narrative can be found in reviewing materials from lobby groups, policy makers or institutional leaders, study abroad staff and many in the research community. The idea is that by living and studying in another country and culture and interacting with local students and people in their community students would become more interculturally competent. This basic notion is

supported by intergroup contact theory, sometimes referred to as contact theory, which was developed by psychologist Gordon Allport (1954).

Intergroup contact theory emerged after World War II, as scholars began to theorize about the nature of intergroup relations and the outcomes that emerge when individuals and groups with diverse backgrounds, beliefs or ideas interact (Pettigrew & Troop, 2008). It was observed that White soldiers in fully segregated units had a far more negative perception of integrated armed forces than soldiers who were already in desegregated units (Brophy, 1946). Observations such as this suggested that in certain conditions intergroup contact could reduce prejudice.

Intergroup contact theory, as advanced by Allport (1954), was based on four principles or conditions for intergroup contact to reduce prejudice. First, there must be equal status between the groups. In this study, the ingroups are the study abroad students; those who have left their home country of study, to participate in the program abroad. The primary outgroups are the local students, teachers, staff, and people that are part of the host culture and environment to which a study abroad student travels. As for equal status, this is difficult to determine. In some cases, the students abroad may not have much, if any, contact with local students, families, the language and culture; and in other cases, significant program design may include ingroup and outgroup contact as a feature of the program.

The second principle is that there are common goals (Allport, 1954). Presumably a program has been advertised to possibly develop a language skill, learn about a new culture, or see an area of study through a different lens and context. In participating in a study abroad program, we assume, though they may articulate it differently, students

going abroad have some interest in intercultural development while away. The third principle is intergroup cooperation. The cooperation in the current study could take the form of local students sharing their apartments and residence rooms with visiting study abroad participants, having students participate in curricular assignments and projects and co-curricular activities on and off-campus. The final principle is support of authorities, laws and customs. This principle is harder to visualize in contemporary study abroad but given that again the ingroup chose to apply for a study abroad to a specific country and location, and that some degree of partnership exists (from direct enrolment to sharing of facilities, classrooms or perhaps faculty members), it could be presumed that there is support from institutional and legal authorities for contact for these groups.

Application of intergroup contact. Though the intergroup contact theory was developed primarily in thinking about relations between estranged or misunderstood groups 'at-home', there seems to be a strong relevance in it for study abroad as well and there are many similarities between the process of intergroup contact and intercultural development. Recall that intercultural competence is defined as the ability to relate and communicate effectively when individuals involved in the interaction do not share the same culture, ethnicity, language, or other common experiences" (FEA, 2018). By increasing knowledge of the outgroup and lowering anxiety, which allows for more intergroup contact, the ingroup or study abroad participant is positioned to have greater empathy with and understanding of the outgroup. This is also aligned with the Engle and Engle model whose classification (reviewed later) suggests that study abroad programs can range from having very superficial observation or situations with minimal outgroup (or host culture) contact to a very integrated, immersed, experiential and reflective

process (2003). The latter situation, the one with far greater and more meaningful intergroup contact, would be the one that presumably leads to more measurable intercultural contact and ultimately competence.

Process of intergroup contact. In thinking about how the process of intergroup contact can lower prejudice and increase understanding and awareness, four principles in action are introduced. The first is that the situation provides a real opportunity for the ingroup to learn about the outgroup. Second is that the ingroup is open to changing their attitudes and behaviours because of contact or immersion with the outgroup. A third is that in bringing together the ingroup and outgroup, there is a possibility of developing affective ties with each other, or alternately even friendships. Finally, there is an ingroup reappraisal where individuals, after participating in the interaction, reassess or re-evaluate their conceptions, ideas or opinions of the outgroup (Pettigrew, 1998).

Intergroup contact and study abroad research. Intergroup contact is a theory that has been used as a framework in previous international student and study abroad research papers. The earliest known use of intergroup theory in study abroad research was by Coehlo (1962) a colleague of Allport. Coehlo focused on the use of stereotypes and how they can lead to attitudinal change from study abroad. He argued that confronting dominant mental images and stereotypes of ‘others can help lead to greater understanding before and during study abroad programs. Kamal and Maruyama (1990) conducted a study using Allport's theory in looking at the experiences of Qatari students studying in the U.S. and their attitudes towards the United States. They found contact with host nationals correlated positively with attitude development, while time spent in the U.S. on its own did not. Other research has explicitly examined study abroad

participants including Stephan and Stephan's (1992) research on American students in a short-term program in Morocco. The researchers found students' contact with host nationals was associated with both decreased anxiety and increased anxiety, depending upon the amount and type of contact. They found contact through school, social and community events reduced the anxiety of the students while abroad, while encountering locals in public, bars or restaurants, in fact increased their anxiety. This supports Allport's theory that not all contact leads to positive outcomes.

More recently, scholars using Allport's theory in research about study abroad and intercultural competence, including Lemmons (2016) and Nam (2011) have questioned the degree to which the four conditions can be measured and met, acknowledging the highly complex social situations of study abroad, which vary greatly from some research that used Intergroup Contact theory in more controlled experimental settings. The degree to which scholars can measure 'contact', in these settings, both frequency and intensity has been questioned by many including those who have researched the theory extensively (Petterigrew & Troop, 2006).

In the present study, intergroup contact theory is not measured directly using the principles and conditions developed by Allport, and later advanced by Pettigrew and Tropp (2008). While the study is unable to verify that the conditions of equal status, common goals, cooperation and support are met, we proceed using proxies for intergroup contact including variables relating to whether the faculty are solely from the home country (or include host country), the language of instruction and the housing arrangements of the students and the presence of intentionally designed intercultural learning opportunities. Instead, this study uses intergroup contact as a broad overarching

theory arguing that being abroad and being in contact with a new culture, language, can aid students in developing interculturally. It is a guiding framework used in many studies included in this dissertation (Lemmons, 2016; Medina-Lopez-Portillo, 2004; Nam, 2011; Savicki & Cooley, 2011). Thus Allport's theory directly informs the first research question about the relationship between study abroad and intercultural competence. Allport's theory is also strongly related to the student and study abroad program characteristics section in research questions two and three.

Validity Framework for Thinking about Study Abroad Research

Validity is a vital, old and consistently evolving concept in empirical research, and one that has gained prominence with the increasing use and dependence on assessments, evaluation and tests in a variety of academic, social and professional settings (Kane, 2006; Messick, 1989). It is "an overall judgment of the degree to which the empirical evidence and theoretical rationales support the adequacy and appropriateness of the interpretations and actions on the bases of the test scores or other modes of assessment" (Messick, 1995, p. 741). Though this and other more current perspectives of validity take a very broad-based holistic approach, traditional views of validity were strongly rooted in a quantitative and positivistic framework (Kane, 2001; Messick 1995). However, as the complexity of the constructs studied, and tests delivered expanded, so too did the theory regarding validity in quantitative research. Supported by work from Cronbach (1988), Messick (1989) is credited with helping to unify and expand the thinking beyond a solely quantitative series of correlational exercises, towards a more constructivist notion of validity. In other words, validity should not be viewed as something with a pre-determined end state but as something "established by the

combined strength of theory and evidence that supports a particular interpretation of a measure" (Porter, 2011, p. 47). Nevertheless, this should not prevent an investigation of some the different and nuanced elements, which are commonly considered to comprise validity. In this dissertation, the meta-analytic approach will analyze a large body of research of existing study abroad research and use theory to make statements about the overall relationship and how different variables are related to changes in intercultural competence in study abroad programs.

For consumers of research, a thoughtful review of the validity evidence presented by researchers is a necessary to make informed conclusions about the quality of the arguments and findings used to make decisions. Validity should be, the standard by which we judge the quality of survey research and from that determine how credible the findings are, how relevant they are to the populations with whom we work and if they present an overarching compelling case for policy, investment and action. Though there are numerous conceptualizations of validity and its composition, as modeled by Vogt (2006), I will focus on four: construct (or measurement) validity, internal validity and external validity and consequential validity.

Construct validity. Building the case for construct validity is a complicated, iterative and potentially lengthy process. Construct validity is the “extent to which the measure is related to the underlying construct” (Groves et al., 2009, p. 50). In developing construct validity, it is common that several related forms of validity are used including content, convergent (or divergent) and concurrent (Vogt, 2007).

Content validity. This refers to how well the content of the test, survey or assessment relates to the content it is intended to measure (Newton & Shaw, 2014). It is

complex in any field, but perhaps more so in social science research that often examines attitudes, opinions and beliefs. Many of the constructs, including intercultural competence, are latent and not directly observable (Creswell, 2012). Therefore, a first step is engaging in a thorough review of the existing literature to understand how other scholars have operationalized, or defined, the construct of interest (McMillan & Schumacher, 2006). Then, a researcher can begin to develop a definition, as well as a set of survey items (called operationalization) that they believe represent their construct. Given the complexity of some constructs they may have sub-constructs, or scales which are comprised of items or survey questions.

For example, Deardorff (2004, 2011, 2017) has argued that the construct of intercultural competence has three primary sub-constructs: knowledge, skills and attitudes. Given the complexity of each of the sub-constructs each is best measured by a series of related items. Having multiple items and sub-constructs, provided they all relate to the primary construct of intercultural competence, helps to minimize construct error (Groves et al., 2009). With these items developed, a researcher can then develop a panel of experts to review the items and assess the degree to which the questions designed or chosen in fact measure the desired construct. This adds to content, and ultimately construct validity as expert panels bring a "broad unique insight on [both the] target populations and the information requested in the survey", which can serve as the basis for additional survey improvements (Ramirez, 2002, p. 1).

Content validity can also be supported through cognitive interviewing (Vogt, 2007). This is a process that focuses on how potential participants make meaning of the questions in the survey, the responses available to them and the overall layout and design

of the study (Dillman, Smyth, & Christian, 2009). By interviewing individuals who have similar characteristics to individuals who could receive the survey, researchers gain valuable insight into how the content of the items in the questions are interpreted and how participants make decisions about which response to choose. This improves content validity by ensuring that the items and responses are clear and that appropriate responses are retrievable by potential survey respondents (Willis, 1999). Content validity matters in this study as the items presented must be focussed on the broader concept of intercultural competence, and not another construct of interest.

Internal consistency. Another way to develop construct validity is to perform statistical tests on the survey item data using measures of internal consistency. A series of individual survey questions (or items) that relate to a broader construct or measure which would be expected to have a correlation pattern. These measures, called alpha values, can be computed for each scale or survey construct using Cronbach's alpha. They provide a measure of the reliability of the items in the scale; or how well they fit together (Vogt, 2007). Typically, alpha values above 0.9 are considered to have very high reliability, while values below 0.6 are considered unreliable and suggest that the items in the construct are not well matched (Field, 2013). High values suggest that individuals are answering questions in a similar pattern, while lower reliability suggests that the responses do not correlate with the latent construct.

The process of developing construct validity can be a long and challenging one. There is no single validity exercise that guarantees validity or suggests invalidity. Researchers should use a combination of these techniques iteratively to build the case that the survey items developed are well designed to measure the construct of interest. For

researchers who choose to use an existing survey instrument, they are not expected to recreate all this validity evidence. They should provide the reader examples of other studies that have used the survey on similar populations, and how the original survey designers, and subsequent research on that survey instrument, build the case for construct validity (Creswell, 2012; McMillan & Schumacher, 2006). This relates back to research question two and three sections on research design characteristics, most notably survey instrument choice. Researchers should present their arguments for choosing a specific survey instrument in their methodology section.

Internal validity. Whereas construct or measurement validity examines the degree to which a survey measures what it purports to measure, internal validity is concerned with the relationship of the observations from the data once the survey is completed (Creswell, 2012). An alternate conception of internal validity is what Campbell and Stanley (1966) referred to as plausible rival hypotheses. The question of 'Can the conclusions the researchers make be attributed to the intervention or program they surveyed people about, or could the observed relationship be attributed to some other, non-measured factor?' There are several categories of threats to internal validity including threats to the subjects, instruments and procedures (Creswell, 2012; Schumacher & McMillan, 2006). However, as all study abroad research reviewed in this dissertation uses a correlational design, meaning there is no random assignment of participants to study abroad or not study abroad, I focus on the four threats most relevant to these designs.

Maturation. This threat refers to changes and development that might occur naturally or without the presence of the intervention (McMillan & Schumacher, 2006). It

is expected that over time, individuals, change, develop and grow on their own. This is what necessitates the use of a comparison group. If the measurements do not show statistically significant differences, the researchers cannot say the changes are due to their intervention, as they may be simple effects of time. In this study, we expect that over time individual may develop intercultural competence over time, however we want to know if that development if greater for those who study abroad than who study at home.

Selection. Threats due to selection effects manifest when those who participate in an intervention or program behave, experience, or are simply different in meaningful ways from those who do not participate in the research (Creswell, 2012). To address this threat, a researcher must make efforts to ensure that the characteristics of the intervention group do not differ in meaningful ways from the comparison group (Vogt, 2007). A pre-post design can help a research control for some of these differences in addition to collecting data about the personal experiences and characteristics of all participants, so that in the analysis phase, these variables could be controlled for (Singleton & Straits, 2010). This too would help the researcher make a stronger claim that measured differences between the two groups in a post-test are related to the intervention and not due to other pre-existing characteristics of the participants. The participants in any individual study abroad research paper could have considerably different personal characteristics. As students typically self-select, or at least self-apply to study abroad, selection concerns are significant in this meta-analysis.

Attrition. A third, highly probable threat to internal validity in correlational research is that of attrition, in which individuals withdraw from a research project before it is completed (Creswell, 2012). When this occurs, drawing conclusions can become

problematic, as those who complete a study abroad program or, the post-evaluation survey, may differ from those who do not (Creswell, 2012). In addition, researchers need to examine if the responses or characteristics of those who withdraw differ significantly from those who completed the study. This threat is very difficult to mitigate against, as it would require researchers to either compare results of a pre-test among those who did and did not complete a post-test, or to follow-up with those who leave the study, to attempt to obtain a post-test result from them. The former provides an estimate of any potential difference between the two groups, while the latter would be difficult to implement and potentially has ethical issues.

History. A final threat to internal validity is history, or changes that occur in the environment that can alter the measurements (Creswell, 2012). These events can occur during a study abroad program or prior to, but the notion is that they can have a meaningful impact on the survey results (Vogt, 2007) In study abroad an event like an act of terrorism, a major shift in the political environment in the host country or even a significant or tragic event occurring to a fellow study abroad participant could reasonably be expected to influence how a group of students scored on the pre-test, the post-test or the magnitude of the difference between the two scores. These are often incidents beyond the researcher's control so in reporting their results, denoting any of these major events occurring before or during the study is important.

External validity. External validity differs from internal validity as its focus is outwards on the ability of researchers to generalize the findings from a sample to a larger population (Creswell, 2012). The ability of a researcher to do this rests on factors relating to the sample characteristics (people), the setting characteristics and the temporal (or

event) characteristics (Cook & Campbell, 1979). As the vast majority of the studies in this dissertation were conducted by researchers employed at American institutions and include samples of U.S students who were primarily studying in Europe, caution should be exercised in generalizing this study to broader populations and geographic locations.

The sample relates to the people in the study and the degree to which their characteristics, behaviours, and possible actions are reflective of those in the larger population (Vogt, 2007). Can research that examines students in college be generalized to those who are in university? Can research that includes only undergraduate students in psychology be applied to those across an entire institution? Or could results of a study on the effectiveness of small classes, conducted with students at a private liberal arts university be generalized to much larger research-intensive institutions? Given that study abroad research never involves a random sample, at best a census of a small population of study abroad participants, generalizability is quite challenging (Salisbury, 2011).

Setting characteristics refer to the conditions of the venue where the research was carried out (Creswell, 2012). In study abroad research, this can refer to where the program took place, and when, or for how long it occurred. As study abroad programs occur worldwide, in developing their conclusions, a researcher needs to consider how broadly and to which contexts the results of their program are generalizable. This threat can be mitigated by including a broad sample in the research that includes participants from a variety of locations. The setting can also be interpreted to include the conditions of the intervention. For example, to what degree can the results of a 2-week study abroad program be generalized to that of a full year academic program?

The final area is the event or temporal characteristics. This is the degree to which what occurred during the intervention could reasonably be expected to occur to other participants (Vogt, 2007). In study abroad research, this can be challenging due to differences in setting, and the wide variety of differences in study abroad programs. Students participating in a study abroad program in the Middle East a few years before 9/11 compared to students studying abroad there today could have a very different experience on a measurement of intercultural competence.

In all the cases above, the researcher could mitigate the threats by using randomized sampling techniques in the design phase. If they are not using randomized sampling, then it is vital to collect data before or after the intervention (study abroad program) which would allow to control for the many differences in study abroad programs and participants (Singleton & Straits, 2010; Vogt, 2007). These are a few efforts researchers can make regardless of the instrument they choose to increase the external validity or generalizability of their research.

Consequential validity. A final consideration with respect to validity extends the definitional statement made by Messick to consider the impact, the consequences both positive and negative, intended and unintended of the results of 'a test' (Messick, 1989). As argued by Keller (1985), "if the research in medicine, agriculture, or business disappeared the consequences would be disastrous. If the research in higher education ended, it would scarcely be missed" (p. 7). Keller suggested that research in higher education is designed more for researchers than those working in colleges and universities. Thus, consequential validity applies to this study and to the instruments in questioning how the measurements and results can be used to make meaningful changes

in programs, services and ultimately experiences and outcomes for students. If the study results are mainly for the benefit of researchers, rather than study abroad staff, program leaders or policy makers, it may help to explain why a narrative of the positive impacts of study abroad appear to be so broadly and widely held, irrespective of the findings of the studies.

Summary. The above is a review of the three forms of validity that inform this study. While validity may seem to apply primarily to questions about research design, I view it in a much larger and broader context of research questions two and three where the individual attributes of studies are examined. Construct validity is examined primarily in the review of existing theoretical models of intercultural competence and in the reviews of the survey instruments and the decision to include to exclude specific studies. Internal validity is the focus in the literature review of individual studies in examining the degree to which the conclusions in the study are supported by data and the design. External validity is used in the data analysis of the current study wherein individual studies will be coded for a host of student, study abroad program. Internal validity issues will also be reviewed throughout the section on research design characteristics. Finally consequential validity is reviewed throughout the results of the dissertation in considering what practical implications the study has for a variety of stakeholders. Validity of study abroad research matters as this is what should inform policymakers and institutional leaders in deciding to expand study abroad opportunities for faculty and staff who design programs and select students to participate.

Astin's Inputs Environment and Output Model

The third framework used in this study is Alexander Astin's Inputs-Environment-Outputs (IEO) model (Astin & Antonio, 2012). This framework was developed at least in part due to some of the challenges identified by Feldman and Newcomb (1969) in understanding the impact of college due to differing research designs and methodologies. In looking at the work of Feldman and Newcomb, Astin found a broad mix of longitudinal, cross-sectional, single versus multi-institution studies, statistical approaches and designs for identifying student-environment interactions. Though Feldman and Newcomb were examining a variety of studies that looked broadly at the impact of college on students, a similar mix of research designs are present in study abroad research as well (as noted by Twombly et al., 2012). Astin hoped that the article would inspire improvements in future research studies to reduce the variety in designs and simplify the synthesis process (1970). Measured by citations, Astin's model has been highly influential in higher education research and program designs. The IEO model has been used as a framework for examining several areas of higher education research which examine teaching styles (Baird, 1973), student faculty interaction (Pascarella, 1980), student retention (Kelly, 1996), student productivity (Hu & Kuh, 2003), and student learning communities (Rocconi, 2011) to name a few. Thus, students (I), enroll and participate in courses and activities at college (E), and which are associated with a range of outcomes (O).

Student inputs. According to Astin, (1970), student inputs are the "skills, talents, aspirations and other potentials for growth and learning that the new student brings with him [them] to college" (p. 225). These inputs can be looked at in two categories. The first are relatively static variables such as sex, race, program of study, and pre-college

preparation. The second can be looked at as what may be captured in a pre-test including intended career path, degree attainment goals, motivation, attitudes, beliefs, and knowledge. Astin also noted that inputs can impact outputs directly and by interacting with the environment (1970).

Within this study, six student inputs are examined. The first is participants' gender. It has been well documented that participation in study abroad is dominated by females (IIE, 2017; Salisbury, Paulsen, & Pascarella, 2010; Stroud, 2010). Whereas the very first study abroad program included only men (Hullihen, 1928) and there was some discussion among the program leads if study abroad was suitable for women, currently more than 2/3rd of study abroad participants are identified as female (IIE, 2017). The second input is student program of study. Originally study abroad was positioned as a mechanism primarily for humanities, language and social sciences students. Current data suggests that from 2011 to 2016, STEM students, as a percentage of all study abroad have increased from 21% to 25%, overtaking both business/commerce students (21%) and social sciences (17%) (IIE, 2017). The percentages of study abroad participants from STEM programs and Social Sciences, have followed divergent trajectories in the past ten years.

The third characteristic is year of study. While the original models of study abroad programs were designed almost exclusively for third year students, contemporary study abroad programs are open to students from all years and levels of study (Hoffa, 2007). Study abroad is also available to graduate students and the most recent statistics indicate that graduate (masters and doctoral degree seeking) students comprise 12% of all study abroad participants in the United States (IIE, 2017).

A fourth characteristic is if the students are considered international or domestic. By this we mean are students, before going abroad, considered domestic (i.e. American citizens attending American universities or Canadians attending Canadian institutions), or international (students from outside the U.S. attending a U.S. institution as an international fee-paying visa student). This characteristic is included as it has been increasingly documented that international students are taking opportunities to study abroad in their home country (Gu, Schweisfurth, & Day, 2010). This could mean a student from China, who is pursuing a full undergraduate degree at Queen's University, becomes a participant in the Queen's-Fudan University student exchange program. Thus, this student is an international student (from the Queen's perspective, who is studying abroad in their home country).

A related fifth characteristic is student self-identified ethnicity. It has been well documented that study abroad participants are disproportionally White compared with the percentage they comprise in higher education (Goldstein & Kim, 2006; Salisbury, Umbach, Paulsen, & Pascarella, 2009; Stroud, 2010). However, the percentage of students studying abroad identifying as White has fallen from 78% to 72% from 2010 to 2016 (IIE, 2017). The final characteristic is previous study, travel or work abroad experience. This variable is included as previous experiences abroad are found to increase students' intent to study abroad in college or university (Goldstein & Kim, 2006; Massey & Burrow, 2012; Salisbury, Umbach, Paulsen, & Pascarella, 2009; Stroud, 2010).

These six student characteristics are examined in the literature review and in this study. Unquestionably this is not an exhaustive list. Other researchers have included additional student-level variables like motivation (Burrow, 2010), pre-college or current

GPA (Salisbury, 2011), student involvement in extra-curricular activities (Salisbury, Umbach, Paulsen, & Pascarella, 2009) attitudes towards intercultural issues (Carlson, Burn, Useem, & Yachimovicz, 1990) or even characteristics of the home institution. However, based on a review of existing literature, data related to these variables are rarely published in study abroad research, and not all studies, even at the dissertation level, include even the six student characteristics listed above.

Study Abroad Program Characteristics

The section above examined some of the most relevant student-level characteristics that can impact both participation and possibly the relationship between study abroad participation and intercultural competence development. This section discusses the college environment or the 'E' in Astin's IEO model, which refers to the elements or unique features of a post-secondary institution that can affect a student. According to Astin (1970a), "they include administrative policies and practices, curriculum, physical plant and facilities, teaching practices, peer associations and other characteristics of the college environment" (p. 225). Environmental factors are elements that distinguish one institution, academic program or research intervention from another, often in subtle ways.

In the proposed study, I am informed by the study abroad classification system developed by Engle and Engle (2003) that outlined seven ways in which a study abroad experience may vary. These variables and the ways that Engle and Engle characterized their differences can be observed in Table 1. The first was the length of the time abroad. This refers to if a student spent a full academic year, a semester, a quarter or less while abroad. Second is target-language competence. This refers to the degree of competency a

student has in a foreign language upon beginning the study abroad program. In research that measures foreign language acquisition, this is typically the pre-test score on a language test (Engle & Engle, 2003). Third is the language used in the academic course work of a study abroad program. Historically, most students went abroad and usually in the language of the host country, there are now many situations around the world where the course language is English, even though this is not the native or primary language in the country (Hoffa, 2007).

The fourth variable is the context of academic work. This refers to the nature of the coursework and whether, or the degree to which, courses are designed only for study abroad students, whether they are mixed or if students are directly enrolled into regular courses with local, or host institution students. The fifth factor is type of student housing wherein students are either in homestays, in housing with their own study abroad group or if they are integrated into housing with local students. The sixth category looks at the degree to which the program provided opportunities for guided/structured cultural interaction and experiential learning. The inclusion of this factor focuses on the issue that direct and meaningful and guided interaction, as opposed to personal travel, may provide a more meaningful intercultural learning opportunity for the participants. The final factor is the degree to which guided reflection of the cultural experience is provided. This is aligned with the sixth factor wherein students participate in an experience but then also have meaningful opportunities to reflect and discuss that experience with an advisor or instructor whether on site, at home or virtually.

Table 1

Engle and Engle Study Abroad Classification System

Program Components	Level One: Study Tour	Level Two: Short-term study	Level Three: Cross-Cultural Contact Program	Level Four: Cross-Cultural Encounter Program	Level Five: Cross-Cultural Immersion Program
Duration	Several days to a few weeks	3 to 8 weeks, summer programs	Semesters	Semester to academic year	Semester to academic year
Entry target-language competence	Elementary to intermediate	Elementary to intermediate	Elementary to intermediate	Pre-advance to advanced	Advanced
Language used in coursework	English	English and target-language	English and target-language	Mostly target-language	Target language in all curricular and co-curricular activities
Academic Work Context	Home institution faculty	In-house or institute for foreign students	Student group or with other international students	In house with student groups	Local norms, partial or direct enrolment
Housing	Collective	Collective and/or homestay	Collective, homestay visit, homestay rental	Homestay rental or integration homestay	Individual integration homestay
Provisions for cultural interaction, experiential learning	None	None	None or limited	Optional participation in occasional integration activities	Required regular participation in cultural integration program, extensive direct cultural contact in service learning, work internship

Guided reflection on cultural experience	None	Orientation program	Orientation program	Orientation program initial and ongoing	Orientation program, mentoring, on-going orientation or course in cross-cultural perspectives, reflective writing and research

These are not the only variables that differentiate a study abroad program. Program cost, number of academic credits, pre-departure/orientation training and the availability and frequency of onsite support are also very important ways in which study abroad programs are differentiated. However, based on my reading of the literature, these details are not commonly shared in published studies.

These four frameworks guided the literature review, the study search, the inclusion and coding process as well as the analytical approach later in the study. I begin with Allport as it is the theory that supports that intercultural competence may be related to characteristics of the participants and the study abroad program. From there I extend to validity as it is a vital part of both Astin's IEO model and Engle & Engle's study abroad classification scheme. Validity is a concern in much of the study abroad literature I have read as a graduate student, most specifically related to research design and generalizability of the existing research. Part of the goal of this dissertation is to understand how well supported the narrative claims of the transformative effects of study abroad are in the literature. This is most specifically related to the 'O' or Outcome in the IEO model of Astin. The IEO model helps to inform both the student characteristics (the Inputs in his model) and the study abroad characteristics (the Environments in his model)

and ultimately the analytical approach used in the meta-analysis. Finally, Engle and Engle's study abroad program classification provides some specific guidance in terms of identifying meaningful ways that study abroad programs may differ.

Chapter 3: Literature Review

A vital step in the meta-analytic process is a thorough review of the ways that key constructs and variables are conceptualized and operationalized, also what the research says about the relationships between them (Card; 2012; Lipsey & Wilson, 2001). This literature review focuses on seven topics. The first is a review of the historical development of student mobility globally and later formal study abroad programs. The second topic is intercultural competence in which I focus on the idea of competence, and the history and development of intercultural competence as a construct. This section also examines how scholars have conceptualized and defined it. The third section moves from theorizing to how intercultural competence has been operationalized, specifically through the development of survey instruments. Fourthly, I review the literature that links the two constructs of interest: study abroad participation and intercultural competence. In this part, I focus on literature that looks at the relationship between these two constructs in its simplest form. The fifth section presents a more nuanced examination of the relationship between study abroad and intercultural development, using Astin's IEO model with review of the literature as it relates to specific student characteristics including gender, program of study, year of study, ethnicity and previous study abroad participation. The last section is centred on environmental factors of study abroad program characteristics identified by Engle and Engle (2003): length of student sojourn, language used in course work, context of academic work, types of student housing, provisions for structured cultural interaction and finally, guided reflection on cultural experience. The goal in reviewing the literature related to both the student and program characteristics is to summarize the existing research to make conclusions and generalizations about whether

these characteristics impact the relationship and the magnitude of the impacts. Throughout this review of existing literature (empirical articles, dissertations, technical reports), I identify the methodological approaches including research design, survey instruments, and analytical approaches. Section seven, the conclusion, focuses on why the narrative approach is a useful, but not sufficient, approach to summarizing the research on these constructs and variables; and argues that a meta-analytic approach is superior to descriptively or narratively summarizing the research on the relationship between study abroad and intercultural development.

History and Development of International Student Mobility

While formal, institution-sponsored study abroad programs have existed for less than 100 years, students have independently sought opportunities to study away from their home country or region for centuries (Dhondt, 2008). The origins of student (and scholar) mobility date back at least 2,000 years ago to institutions like the University of Taxila in Pakistan (Hoffa, 2007). Taxila provided instruction in Greek and Persian and drew scholars from across the region including Alexander the Great (Brickman, 1961). Centuries later, the University of Nalanda in India became a well-known and highly regarded institution and attracted numerous scholars from China from about 450 to 1100 AD (Brickman, 1961). Much of what is known about Nalanda as a place of higher learning is a result of the documentation from these Chinese scholars (Hoffa, 2007). Around 400-500 AD, cities and centers like Alexandria, Antioch, Rhodes, Athens and Rome became very popular sites to study for foreign students and scholars due to their extensive library holdings. Individuals from across Europe, Asia and the Middle East flocked to these locations to learn from the vast collections of literature, readings and

knowledge, maintained in these cities, but also to learn the culture of the people living in those cities (Hoffa, 2007). Rome became such a powerful draw that the Emperor felt compelled to issue and order all visiting students and scholars to present their passports, residence credentials, and field of studies to the local Registry. If anyone was found engaged in behaviour "discreditable to the interests of learning", the Registry warned they would be removed from the city and sent home immediately (Capes, 1922). The attractiveness of these locations was a combination of the institutions themselves and the fact that these cities were major economic and cultural centers, making them like many of the most popular study abroad destinations today.

During the medieval, post Roman-empire period, mobility slowed considerably. However, records of students from the Nordic countries travelling to study on the continent have been found dating back to the 12th century (Dhondt, 2008). These students were primarily from elite families associated or affiliated with the Church and later became part of the ruling classes themselves. The proliferation of Latin as the language of instruction throughout much of Europe enabled Nordic students to continue studying on the continent for several hundred years (Dhondt, 2008). During this time, and until the Renaissance, a smaller community of scholars and learners continued to traverse beyond national borders for several centuries.

The earliest European universities became major hubs during the Renaissance for mobile students and scholars including those in Paris, Oxford, Bologna, Salerno and Montpellier. Though European scholars and students comprised most of the names on these surviving student lists, there also existed significant numbers of scholars and students from the Middle East, North Africa, and the West Indies (Hoffa, 2007). During

the Renaissance, the presence of foreign students in some centers had grown to such a degree that some began to question if the local institutions were catering too much to these visiting students and scholars. Many local and international scholars were concerned about how many students failed to return to their homes after completing studies ‘abroad’, highlighting an ongoing issue of brain-drain/brain-gain. Others highlighted that some students were earning doctoral degrees despite not having any knowledge of Latin at all. These concerns about the language of study by visiting and foreign students are akin to contemporary issues of the prevalence of English as the language of instruction for both international students and in study abroad programs, even where the host language of the country is not English (Wolfe, 2007). It could also be likened to growing importance of international students, though primarily from China, being recruited to study for their full degrees in Canada, the United States and Australia (Sá & Sabzalieva, 2018).

Another concern was the sheer volume of visiting students in some of these cities, leading to locals questioning if their city was losing its uniqueness, and if the visiting students were in fact learning about the city (Hoffa, 2007). These concerns have relevance today as some of the more critical voices in study abroad (see Woolf, 2007; Grunzweig & Reinhart, 2002) have noted that some European study abroad centres like Florence and Barcelona are flooded with visiting study abroad students; a problem that becomes exacerbated in the summers with the additional influx of college-aged tourists from around the world.

The Grand Tours of Europe were the next significant movement in international (Western) student mobility emerging after the Renaissance (Brennan, 2004). The Grand

Tours differed from some of the previous examples of student mobility as the Tours did not involve study and typically occurred after finishing university or other higher training. The leading political, religious and commercial families saw the need to have their heirs (sons) carry on after them and saw great value in having their children spend time abroad (Hoffa, 2007). These tours involved both relationship-building with other powerful families but also a large focus and interest in greater cultural exposure and enlightenment. The Grand Tours were viewed as liberating socially, academically, personally and aesthetically (Buzzard, 1993).

The decline of the Grand Tours coincided with the development of the railway system in Europe. While on one hand it made travel across European countries far more convenient, it also made access to these previously 'exotic' locations available to a broader segment of the population (Dhondt, 2008). Thus, the Grand Tours were no longer something that could be accessed only by the elites; members of the middle class could visit and learn in these cities too. Currently, university students may take a gap year during (or even before) their studies to travel, pursue work or other opportunities abroad that may or may not be related to their area of study or career interests. It is argued that these gap years provide young people an opportunity to mature, develop personally, become more self-sufficient, and develop a greater or broader understanding of other parts of the world (Lyons, Hanley, Wearing & Neil, 2012).

The examples above involving the movement of students and scholars provide a foundation and a backdrop to a centuries-long history of student mobility that has existed around the world. Still these examples are not typical of what is considered study abroad in the current study, as many did not involve formal academic coursework or training for

the purposes of degree attainment. What they did include was a strong focus on movement to a locale wherein students could develop cultural knowledge and build social capital. A key element of study abroad is being hosted at an institution or program outside of their current country of study and then returning to continue their studies, or graduate, from their home institutions.

History of study abroad in the United States. The origins of what would eventually evolve into the first sponsored study abroad programs by a university may have started at Indiana University in the 1870s. During the summer breaks, several faculty members developed, hosted and supervised groups of students who participated in a series of "summer tramps" through various parts of Europe (Indiana University, 1879). These tramps began as site-seeing and cultural tours, like the Grand Tours of Europe. However, sometimes instruction and conversation between faculty and students was in the language of the host country; typically, French, German or Italian. These programs also had a thematic focus such as natural history, languages and culture and the programs evolved to a degree that the university eventually began giving credit for them in the 1890s (Indiana University, 1879).

Though the IU summer tramps may have been the first to offer credit for course work or experiences abroad, formalized study abroad programs in the United States did not emerge until after the First World War. The study abroad programs that began to appear in the 1920s fell into one of three categories. The first was the Junior Year Abroad (JYA), which involved a student being abroad for the entire academic year with a focus on learning the language, history and culture of the host country (Hoffa, 2007). This is best characterized by the program at the University of Delaware which created the first

formal, university-sponsored, study abroad program in 1923. The university sent eight, third-year students on a year-long program to the Sorbonne in Paris, France (Hullihen, 1928). The program was designed to provide an opportunity for students to immerse themselves in full-time study at another university so that they could both improve their language skills and understand another culture (Hullihen, 1928). Initiated by Professor Raymond Kirkbride, and partially funded by the DuPont family, the Sorbonne program was successful enough to be replicated by eight other U.S. universities by the end of 1920s (Abrams & Hatch, 1960).

The second model was the faculty-led study tour. These programs involved American faculty members guiding students to specific locals (often in Europe), teaching courses in English, though possibly with some language instruction in the host language. New York University was one institution that hosted courses abroad, recognizing the growing role of foreign affairs in a post WWI world. It developed several opportunities for students, with faculty members from the United States leading courses on the growing field of interest in international political affairs (Hoffa, 2007). This faculty led model is evident in what is now known as Semester-at-Sea. An early version of Semester-at-Sea, called the floating university world cruise, began in 1926 and was restarted later in the mid-1930s (Hoffa, 2007). This program was not delivered by a college or university but by an organization called the University Tour Association. This organization had numerous university presidents on its advisory board, who were recruited to help increase the credibility of their tours and to gain insight on how to design the program so that it would be accepted for credit by institutions (Hoffa, 2007). These faculty led tours differ from many of the junior year abroad programs in that the students are in courses and

programs designed especially for them; there is often little to no coursework with local students.

The third model is short-term, often campus-sponsored, study abroad programs. Initially these were not-for-credit, more akin to the IU Tramps, or Grand Tours, as they were not entirely, or perhaps even primarily associated with post-secondary education but with tourism. However, given the cost of the programs, they eventually evolved into credit-bearing opportunities making students and their families more willing to invest the money to participate in these unique educational programs. These programs occurred outside the traditional academic year, occurring between semesters and often during the summer, in a more intensive format (Hoffa, 2007). Georgetown offered a series of World Issues courses, which had prerequisite courses to be taken at home, followed by the tour abroad in Latin America (Hoffa, 2007). These tours allowed students a chance to become immersed, for a short time, in the local culture and investigate how the theoretical aspects of law, economics, history and politics were lived in the host country. IU, well known for their summer tramps offered a unique summer program in World Music in Munich, Germany (Hoffa, 2007). Students had the opportunity to be instructed by local musicians and faculty, in addition to their home faculty, as well as attend symphony and opera performances during their summer program. Short-term programs are now the primary study abroad design in U.S. post-secondary education with nearly 60% of all programs 8 weeks or less, often during the summer or between semesters (IIE, 2018).

In the U.S., federal support for study abroad identifies global competitiveness, national security and international leadership as reasons for enhancing study abroad opportunities for university students (Commission on the Abraham Lincoln Study Abroad

Fellowship Program, 2005; Department of Education, 2012). The initial Simon Study Abroad Act (SSAA) was introduced to the U.S. Congress in 2006 to help meet these governmental goals. The overarching vision of the SSAA is to increase the number of American students participating in study abroad programs from 250,000 to 1,000,000 annually. However, current participation is approximately 330,000 (IIE, 2017). In addition, the SSAA aims to help increase the quality of study abroad experiences, create more certainty for credit transfer, and enhance the diversity of programs and variety of the students participating in study abroad programs. Funding to support these goals will be earmarked for institutional grants to support the development of unique academic programs, programs in non-traditional locations, and/or those serving non-traditional populations of students. More recently, the United States has added an additional program goal of having 100,000 American students study abroad in China each year, though not all at the post-secondary level (U.S.-Strong, 2018). While national goals for study abroad are not uncommon (e.g. Canada's unofficial goal is 50,000 abroad annually and Brazil's is 100,000), this is one of the first large-scale initiatives to specifically target study abroad participation in a specific country.

Development of study abroad in Canada. While the history and roots of student mobility and study abroad in the United States and Europe have some historical paper trail, tracing the development of these programs and activities in Canada is problematic due to a lack of scholarship (Bond & Lemmanson, 1999). Lewis Perinbaum, who helped establish the first on-campus study abroad offices, advanced the precursors to the current university-sponsored study abroad programs in the 1950s and 1960s (Bond & Lemmanson, 1999). These offices served to facilitate participation in, and the

development of, what is now the Canadian University Service Organization (CUSO). CUSO serves to place Canadian scholars and technical experts in strategically important emerging nations. Now, more than 50 years after Perinbaum's efforts, nearly all universities in Canada, and many individual faculties, have established study abroad and exchange offices.

Development of ERASMUS in Europe. Within Europe, though student mobility has existed for several centuries, the formalization of relationships between institutions and nation-states is much more recent. Europe is unique for its transnational, high-profile intergovernmental student mobility (their preferred term) previously known as the ERASMUS programme (European Regional Action Student Mobility). ERASMUS was developed by the European Union in 1987 to increase the volume and improve the ease of student mobility within Europe (European Union, 2011; Teichler, 2017). While the ERASMUS program is approved and funded by the member countries, individual universities determine participation. The budget provided annually in scholarships and awards for students, staff and faculty participating in Erasmus sponsored exchange programs is more than $CDN 500,000,000. This funding supported the exchange of more than 330,000 students among EU member countries in 2016/2017 (European Union, 2017). The European model for supporting and developing student mobility is strongly promoted at the trans-national level and identifies student development and institutional relationship building as its priorities. It was created to foster greater ties among European citizens, institutions and countries (Mitchell, 2012).

Critical Study Abroad. The section above discusses study abroad in a largely uncritical fashion. The definition used in the study ignores the fact that in Western

countries post-secondary educational attainment is only now approaching 50% (World Bank, 2019). And of those who are able to be in post-secondary education, only a very small percentage of those students participate in study abroad annually. These students (details shown in table 9) are not representative of the overall population of post-secondary students. Despite this lack of representative, discussions of study abroad and study abroad participants often ignore the cumulative privilege of "race, nationality, education, mobility, class" (Zemach-Bersin, 2008, p. 2) and I would add ability. Twombly et al. (2012) summarize the literature on why a very large percentage of students do not express interest in, let alone intent to, study abroad by citing four factors; socioeconomic status; a paucity of information about opportunities and requirements; involvement in related activities, attitudes and motivation. It is also rarely acknowledged the degree to which study abroad, an activity in which just 2-3 percent of post-secondary students participate in, produces, enhances or reproduces social capital (Reilly & Senders, 2009).

There are many forms of study abroad programs, but two of the most common are the faculty (or third party) led program or the bilateral/consortia exchange. The former involves faculty or administrators creating a program abroad for a group of, usually, American students, while the later has students from two or more countries effectively switching places in their institutions. The faculty-led program is designed to have the home country students learn and develop without specific attention to experience of the home country or culture. Though a thoughtfully designed program would make efforts to provide meaningful interaction. In contrast, the bilateral agreement is predicated on students from different countries having the opportunity to experience each others

institution, country and culture. At the same time it can be argued that bilateral agreements are more likely to exist between Western or English-speaking countries, whereas the faculty-led programs could be developed in nearly any country or local worldwide. Finally, when students do study, at least partially, in another language while abroad, these are most often the colonial languages of French, German and Spanish (Reilly & Senders, 2009). And there is a growing trend of students taking courses primarily or exclusively in English while abroad, even in countries in which it is not the primary language (Hoffa, 2007; IIE, 2017). While lobby groups, professional associations and institutional leaders rarely, if ever, acknowledge some of these critiques, a number of researchers continue to challenge the dominant narrative, the purpose, and the beneficiaries of study abroad programs and programming (see Gore, 2005; Woolf 2008, Zemach-Bersin, 2012 for additional research).

Conclusion. This first section of this review has traced the history of student mobility around the world, highlighting that individual students or scholars have long sought out opportunities abroad for academic, professional and intercultural competence. In the last century, this evolved into formal university-developed, sponsored and recognized programs and opportunities which have facilitated opportunities for a greater number of students than ever, and with plans for even more students to study abroad in the future. The section concluded with a brief, critical look at study abroad, one through a lens less discussed by many stakeholders.

Measuring and Defining Intercultural Competence

Though a key element of the definition of study abroad is that the students can earn credits towards their degree, diploma and program of study, much of the rhetoric

surrounding study abroad and why students should participate centers on the potential for developing greater intercultural competence. The definition used in this study again is "the ability to relate and communicate effectively when individuals involved in the interaction do not share the same culture, ethnicity, language, or other common experiences" (FEA, 2011, p. 44). The next section reviews the development of the construct itself, but also the two key elements of the construct: 'competence' and 'intercultural'.

Defining competence as a construct. The idea of competence and competency among post-secondary students and graduates is a topic of much discussion in popular media, scholarly research and policy circles. But the idea of competence in education can be traced back to Taylorism in the 19th century and notions of scientific management (Bowden & Marton, 1998). In the workplace, Taylorism involved analyzing the individuals and required tasks to identify ways to measure and ultimately improve efficiency. In education circles, competence can be seen in the work of Benjamin Bloom. Bloom viewed competence as "directly observable behaviours which can be reliably recorded as either present or absent" (Bloom, Hastings & Maudas, 1971, p. 28). His taxonomy stated objectives, or outcomes, in terms of students' ability, or inability, to state, list, record, match, describe.

Bowden and Martin (1998) expanded upon Bloom's behaviourist perspective in three ways. The first is additive, which extends performance with an additional test of knowledge. The second is integrative, wherein performance and knowledge would be integrated. The third addition was a holistic notion of competence. This includes the previous iterations but adds an individual's self-perception and view of their own

performance in knowledge. More recently, perceptions of competence have been extended to suggest that competence is influenced by people, goals and context and that it should be examined within a constructivist more than an objectivist paradigm (Stoof, Marten, van Marrienboer, & Bastiaens, 2002). This inclusion is aligned with Astin's IEO model, which posits that measuring student outcomes should be done in consideration of the pre-existing student (the people) and their college/program of study (context) characteristics. More recent characterizations of competence have focused on "context specific dispositions which are acquired, and which are needed to cope successfully with domain-specific situations and tasks (Blomeke, Zlatkin-Troitschanskaia, Kuhn, & Fege, 2013, p. 3). That approach fits well with the definition of intercultural competence and the rationale for encouraging students to participate in study abroad programs.

Moving from generic descriptions of competence to the area of intercultural competence and incorporating all the ideas above, the focus moves to situation-specific skills, attitudes and knowledge, but also the behavioural capacity and ability to integrate diverse experiences into a current situation. This allows for a real situation where an individual may be (or perceive themselves as being) competent in one situation, but less so in another with a different set of circumstances. Thus competency, like validity, is something that is strived for, but unlikely to be ever fully met (Spitzberg, 2007; Messick 1989). How researchers and scholars attempt to measure intercultural competence is the subject of the next section.

History of intercultural competence in (and outside of) study abroad. While the phrase "intercultural competence" has only arisen in the past 30 years, interest in the potential of developing cultural skills and learning about different cultures, dates to the

Summer Tramps hosted by Indiana University. The participants and program directors of the first Junior Year Abroad programs, while primarily focused on academic coursework and language acquisition, noted the major benefits of living and learning about another culture while abroad (Hullihen, 1928). Hullihen noted that students returning from Paris had "contact with an atmosphere of cultural and aesthetic ideals quite new [to them]" (p. 30). College officers associated with the University of Delaware and other early study abroad programs also commented on the growth of their students' knowledge and understanding of the local culture (Hullihen, 1928). These were early recognitions that the benefits of study abroad went beyond academic credit, which would be the knowledge and possibly the attitude aspects of competence. Moreover, it is not clear that these students were developing what might currently be construed as intercultural competence, as they were primarily being exposed to rather than being deeply engaged in the local culture while abroad.

Other programs during these early years of study abroad also noted how their experience abroad helped them reflect on what it meant to be an American, which relates to the attitude element of competence, in addition to broader knowledge of the cultural context of the language they were learning. One example was the Montclair Teachers College program, which was oriented around training and developing the foreign language skills of teacher candidates abroad to make them more capable of teaching when they entered the school system (Haddis, 2000 from Hoffa, 2007). Despite the recognition of the potential cultural learning opportunities that these study abroad programs provided, the degree of immersion and intentional focus on intercultural learning was limited by the lack of host nation language skills of the study abroad

students (Abrams, 1967). In fact, the students' cultural experience was impacted as much by the families these students roomed with as their peers. In addition, the experiences abroad were limited to just a handful of countries including Spain, France, Germany and Italy (Hoffa, 2007). Even though the number and profile of study abroad participants have expanded greatly since the 1930s, concerns remain and the factors that inhibit intercultural learning, and even interest in study abroad participation (Vande Berg, 2009; Woolf, 2007).

Except for a few programs in Latin America and Canada, WWII and its aftermath effectively put study abroad programs on pause for nearly a decade. This cessation of study abroad programs by colleges and universities meant that there were several cohorts of college and university students that went through their post-secondary experience not knowing about, or having the opportunity to, study abroad. However, the post WWII period was also one in which interest in international politics, economics and area studies among students at home grew significantly (Hoffa, 2007).

The post-war period was also characterized by a growing interest in the experiences of how individuals from the West functioned in their work abroad (Bennett, 2010; Sinicrope, Norris, & Watanabe, 2007). Researchers developed an interest in learning about the challenges that individuals working in organizations like the Peace Corps or Foreign Service had in working and communicating in their international roles (Gordon, 1967; Guthrie & Zektick, 1967; Mischel, 1965). This work focused on identifying the qualities that might predict employees' or volunteers' success in working abroad. In addition to a growing interest in the intercultural elements of overseas work, the postwar period also saw renewed interest and growth in study abroad program

participation (Hoffa, 2007; Abrams & Hatch, 1960). In the 1960s, as participation in study abroad among American students began to grow again, "the furthering of international understanding," rather than purely academic coursework, became a more significant impetus for study abroad (Abrams, 1960, p, 9). Other scholars saw the potential for greater 'international understanding' and a better understanding of the life and culture of the United States both among students and between nation states from greater participation in study abroad (Gullahorn & Gullahorn, 1960).

Intercultural Competence Post World War II

Research after the postwar period brought forth some of the first empirical and theoretical efforts to study and make meaning of the cultural and intercultural processes in living abroad and interacting with individuals from and in diverse cultural contexts. It was in this period that Allport, building off earlier work by Brophy (1946) and Williams (1947), began his work on intergroup contact as a mechanism for reducing prejudice. A major advance in thinking about culture was developed by Hall (1959), who described cultures in terms of high and low context cultures in communication styles. According to Hall (1959), high context cultures tend to use less explicit or formal approaches to communication and more internalized understandings of what communication is occurring. High context cultures are characterized by a focus on longer term relationships and a strong focus on face-to-face relationships with persons of authority. By contrast, low context cultures are more rule-oriented and individuals operate by external rules. Low context cultures are also more likely to have a focus on shorter duration of contacts and relationships and exist in situations where knowledge is more transferable and highly task-specific. These theories were welcomed by those involved in international work as

they provided some practical guidelines, though not rules, for understanding how to communicate and interact with individuals from other cultures (Bennett, 2010). Hall's major contribution here in terms of study abroad was providing a framework for thinking about how communication occurs within cultures, and strategies to bridge gaps when communicating with someone from a culture different than one's own.

While Halls' work focused primarily on communication between individuals, the work of Lysgaard (1955) introduced ideas of how individuals who live, work or study in another cultural context adapt and change to these new surroundings over time. Lysgaard's (1960) work, built on the ideas of culture shock first developed by Oberg (1953), who described the causes, cures and recovery for individuals in new cultures. Lysgaard (1955) developed the U-Curve, which suggests that individuals in new cultural contexts will begin in a honeymoon phase, then encounter a crisis period, followed by recovery, then readjustment and ultimately a re-entry to the original cultural context. Later, Gullahorn and Gullahorn extended the U-curve hypothesis to a W-curve in 1963. Their work added a series of stages that happen to individuals once they re-enter their home culture, arguing that that experience is also one that presents some shocks at home not just in the 'Alien Social System'. Their additional stages begin with a honeymoon at home phase, followed by crisis at home (re-entry shock), then recovery at home and (re)adjustment. Both models hold a place in intercultural training sessions as well as study abroad pre-departure and re-entry sessions as guides to help students prepare for and anticipate the changes and experiences that they may experience abroad. While numerous studies attested to the validity of the U- and W- curve hypothesis models (Becker, 1968, Davis, 1963; 1971), more recent empirical work did not substantiate those

claims. For examples, Black and Mendehal (1990) found that while two-thirds of studies support the theory, few were based on, or provided, evidence of the statistical analysis to support their arguments. Another examination of the U-curve by Ward, Okura, Kennedy, & Kojima (1998), also found that the U-curve hypothesis could not be supported empirically. Though both the high and low context communication styles of Hall and the U and W curve theories have considerable intuitive appeal and remain staples of many study abroad pre-departure training sessions, the empirical evidence suggests that use of these models should be reconsidered. This is not to say that culture shock as a phenomenon does not exist, only that it may not follow the relatively smooth patterns that have been suggested. Instead, it has been argued that researchers should expand their focus in understanding the cultural learning and adaptation process to include individual motivation, learning processes and environments (Black & Mendenhall, 1990). This recommendation foreshadowed the suggestions Engle and Engle (2003) advanced about how study abroad program characteristics may affect intercultural development.

Additional collaborative ventures in the 1960s and 1970s included the Intercultural Communication Workshop, which was designed to advance the ideas of intercultural training, theory and practice (Bennett, 2010). The ICW was a series of intensive courses, between 15 and 30 hours in length that included a mix of U.S. citizens and individuals from another culture. Trained facilitators would lead the groups through discussions and dyad interactions and exercises (Bennett, 2010). From 1966, when the first formal ICW program was hosted by Hoopes and Rhinesmith, to 1973, at least 11,000 visiting students and 7,500 American students participated. Evaluations of these programs found mixed support for their effectiveness. Some found the workshops

improved cross-cultural friendships and interactions but found no relationship to (positive) changes to cross-cultural attitudes (Hammer, Gudykunst, & Wiseman, 1979). Hammer (1984) found no significant differences in six measures of intercultural competence between participants and non-participants in a related study. However, Hammer did note that a reliance on t-tests, which produces results in terms of significance, might be masking a meaningful effect due to how sensitive the techniques are to small samples,

Despite the mixed results of the ICWs, they were highly influential in advancing future research and practice in intercultural communication. Bennett (1986) summarized some of the learnings from the ICWs in noting that cultural patterns should be differentiated from cultural labels. By this he means that understanding the cultural patterns is the first step to the intercultural learning process, which can lead to the development of specific cultural strategies. The ICWs supported the notion that culture-general frameworks are necessary and useful in that they help guide but not directly influence individual's perceptions. Being able to identify some language, non-verbal, communication, cognitive approaches and cultural value differences allowed participants to identify and adapt to other, previously unidentified clues (Bennett, 1986).

These were some of the key early theorizations and developments in the move from an almost happenstance surprise that study abroad had possible cultural benefits, to theorizing about intercultural communication approaches and adjustments that individuals encounter while in new cultural contexts. In addition, intercultural training workshops across the United States, laid a foundation for future scholars of intercultural

communication and for increasing complexity in the way that various elements of intercultural competence have been theorized and operationalized.

Contemporary models of intercultural development

Throughout the 1970s, 80s and 90s, the theoretical approaches and models about intercultural competence grew significantly, both in nominal terms and in complexity of approach (Deardorff, 2006; Sinicrope, Norris & Watanabe, 2007). An increasing number of approaches, studies and research tools were developed to measure and compare the intercultural competence of different populations in varying academic and professional settings. The next section provides a review of some of the various approaches with a focus on the key constructs identified by researchers in these models.

Hanvey's Global Perspective. One of the first attempts to operationalize constructs related to ICC was by Hanvey (1976) who advocated for a more intentional approach to helping students develop what he called a global perspective. By global perspective, Hanvey was describing the "learning which enhanced the individual's ability to understand his or her condition in the community and the world and improves the ability to make effective judgements" (p. 1). In this model, a global perspective had five dimensions, which included perspective consciousness, state of the planet awareness, cross-cultural awareness, systematic awareness and options for participation. Returning to the discussion of competence, Hanvey's model focuses on attitudes, not knowledge or information about other cultures or cultural practices. His work was unique in that it focuses specifically on the importance of enhancing the global perspectives of students in the K-12 and post-secondary system. Hanvey's work led to the development of the Global Perspectives Survey, which will be discussed later.

***Ruben & Kealy behavioural approach*.** Appearing at nearly the same time as Hanvey's largely personality and attitudinal approach was a more behavioural model advanced by Ruben and Kealy (Ruben, 1976; Ruben & Kealy, 1979). They focussed on the gaps between what individuals believe to be intercultural competence and what they do in an intercultural situation. There are seven dimensions in this model: displaying respect, interaction posture, orientation to knowledge, empathy, self-oriented role behaviour, interaction management and finally tolerance for ambiguity. For assessment purposes, Ruben created observational procedures and rating scales and others have used his model to create assessment tools as well (see Fantini, 2006; Koester & Olebe, 1988; Ruben, 1976; Sinicrope et al., 2007).

Byram model of ICC. The two models specified above originate out of the U.S. from a largely mono-cultural perspective. The focus on language is part of what makes Byram's model unique. He was interested in the context of individuals who teach in foreign cultures and individuals who teach foreign languages. The Byram model uses five dimensions of intercultural competence both latent and observable: attitude, knowledge, skills of interpreting and relating, skills of discovery and finally critical cultural awareness (Byram, 1997). This model clearly identifies the attitudes, knowledge and skills components as being the primary components of competence.

***Bennett Developmental Model of Intercultural Sensitivity*.** Milton Bennett has developed what is now one of the most commonly cited theoretical models of intercultural competence, especially as it relates to study abroad. The model is influenced by his years of experience working with the Intercultural Communication Workshops. The Developmental Model of Intercultural Sensitivity (DMIS) is a developmental model

whereas the previous ones were compositional (made of components) or co-relational (factors are related to each other). The developmental aspect refers to stages that individuals grow and develop through. The underlying assumption of the DMIS is that as "one's experience of cultural difference becomes more complex and sophisticated, one's potential competence in intercultural relations increases" (Hammer et al., 2003, p. 423). Thus, individuals move from monocultural to more differentiated multifaceted worldviews, which are characterized by three ethnocentric stages: denial, defense, and minimization; and three ethnorelative stages: acceptance, adaptation and integration (Bennett, 1986). This developmental approach is like Perry's (1970) theory of intellectual development where individuals advance from dualism to multiplicity, to relativistic, and ultimately committed relativistic positions. However, in terms of other intercultural theories, this model has more in common with the U and W-curves, which are also developmental as they posit that intercultural development or communication involves a movement and adjustment through a series of stages. The DMIS is the basis for one of the most widely used survey instruments in measuring ICC, the intercultural development inventory (IDI).

King and Baxter Magolda Model of Intercultural Maturity. Like the DMIS, the King and Baxter-Magolda Model of Intercultural Maturity (2005) is also a developmental approach. It theorizes a three-level, three-factor model of intercultural maturity. The Model of Intercultural Maturity suggests that individuals move from initial to intermediate and finally mature levels of intercultural maturity by developing their cognitive, intrapersonal and interpersonal skills. Like the DMIS, the model posits that development occurs through ongoing study, observation, and interaction with

intercultural situations. As with many of the previous theoretical models, the Model of Intercultural Maturity forms the basis for a survey instrument, The Global Perspectives Inventory that is widely used in study abroad research but also to examine intercultural development in higher education more broadly.

Critical Intercultural Competence. The preceding sections on intercultural framework and instrumentation are presented largely from a positivist and uncritical position. Before proceeding, three critiques should be examined. First, all of the frameworks and the instruments reviewed and cited above, are produced by Western scholars for use in Western contexts. Wahyudi (2016) questions why so few scholars from outside Western circles have been involved or are consulted in the development of the frameworks and instruments and why so few are included in the major publications about the construct. He hypothesizes that the omission of non-Western scholars may partially explain why the models and instruments center intercultural competence on the individual, than a more group or interpersonal approach as found in many Asian contexts (Yum, 1994). Wahyudi argues that intercultural scholars and researchers should place a greater focus on the postnational (effects of globalization), the postmodern (status of knowledge and concepts of self) and the postcolonial (the colonial characteristics of 'English') in their work. This might reorient some of frameworks and instruments towards a more dynamic conception of intercultural competence. These frameworks would address power relations between individuals and groups from different countries. And the power relations themselves are situational. They evolve and change depending on the identities of those involved in intercultural situations and less individualistic (Williams, 2006).

A second critique questions the purpose of developing or enhancing intercultural competence. A foundational document to enhance U.S participation in study abroad was the Commission on the Lincoln Study Abroad Fellowship Program (2005). This document explicitly tied the development of greater cultural competence through study abroad participation to a national response to globalization, economic competitiveness as well as to secure national security, especially in a post 9/11 environment. It, like many other professional association and lobby group reports and publications on study abroad, did not recognize intercultural competence as a process for valuing the local, or the complexities in that learning. Instead the focus was on static, reductionist and idealized notions of culture and learning (Reilly & Senders, 2009).

A final critique, related to those above and that of study abroad more broadly, is centred on the idea of who gains from the development of intercultural competence. In a review of major professional associations approaches to internationalization, Buckner and Stein (2019) ask

> To what extent are the outcomes of internationalization framed in terms of "acquiring" knowledges about places and peoples, rather than "deconstructing" assumed knowledge, and "opening up" new possibilities for relating to ourselves and the world in ways that account for our differences and interdependencies? (p. 14)

Using their framing, ICC is often positioned as something to be developed and gained while abroad for 'use' back home; this privileges both the individual and the (typically) Western country the student is coming from. Tervalon and Murray-Garcia argue for an approach that embraces 'cultural humility' (which they consider 'a commitment and active engagement in a lifelong process that individuals enter into on an ongoing basis with patients, communities, colleagues, and with themselves' (1998, p. 118). They argue

that a focus on humility returns critical reflection to cultural interactions, mitigates and makes students more aware of power imbalances and does not patronize the host country, culture and environment (Murray-Garcia & Tervalon, 2017). This approach aligns well with Allport's (1954) principles of equal status among groups and intergroup cooperation.

Summary

The above models represent a handful of the dozens, or more, conceptual and theoretical models of intercultural competence. They are presented to highlight the various approaches to conceptualizing intercultural competence. They can be compositional (see Deardorff, 2006), co-orientation (Byram, 1997) and developmental (King & Baxter-Magolda, 2005; Bennett, 1993, Gullahorn & Gullahorn, 1963). This highlights a sample of the various approaches and underscores how many conceptual approaches exist. These models also underpin many of the survey instruments that now form the basis of much of the quantitative research on intercultural competence in study abroad. Thus, understanding the theory that the survey instruments are based upon is important as it provides information about the ways that researchers conceptualize the intercultural learning process, and the primary elements of intercultural competence.

As the field and volume of scholars and scholarship has grown, individuals have periodically called on the need for "conceptual clarity" (Ruben, 1989, p, 234) and the "need for a clearer definition of the concept" (Kuanda, 2004, p. 10). Despite the plethora of approaches, only a few scholars have attempted to narratively or empirically review the landscape of these models. One of the most significant attempts to unify these approaches was conducted by Deardorff (2006) in her doctoral thesis. She sought to

develop a definition and identify what scholars and institutional leaders believed were the ideal assessment approaches to measuring intercultural competence. Deardorff used a 3-round Delphi technique where the scholarly experts developed and submitted definitions of ICC, which then went through three rounds of revision. This research found that the definition found to be most relevant to institutions' internationalization strategies was one based on Byram's (1997) definition that ICC was "knowledge of others; knowledge of self; skills to interpret and relate; skills to discover and/or to interact; valuing others' values, beliefs, and behaviors; and relativizing one's self. Linguistic competence plays a key role" (p. 34). The second highest rated definition was Lambert's (1994) "Five components: World knowledge, foreign language proficiency, cultural empathy, approval of foreign people and cultures, ability to practice one's profession in an international setting" (as cited in Deardorff, 2004, p. 230). These findings did not constitute a consensus in the field but highlighted that the themes of knowledge, skills and attitudes are apparent in two of the most commonly cited definitions. Deardorff's research showed that there is some broad agreement in terms of the conceptual elements of intercultural competence, but at the same time, there was considerable diversity in the operationalized elements or key constructs of it. The next section illustrates some of the approaches scholars have used to operationalize intercultural competence for building survey assessment tools.

Operationalization of Intercultural Competence

The previous sections highlighted some of the attempts to theorize and conceptualize models of intercultural competence. Like most latent constructs, there is no commonly agreed upon definition let alone term. Scholars have catalogued at least a

dozen terms including “intercultural competence, intercultural communicative competence, global competence, global citizenship, multicultural competence, cultural fluency, communicative competence, cultural competence, intercultural sensitivity, cross-cultural awareness, cultural intelligence, cultural literacy, cross-cultural capability, (Deardorff & Jones, 2012, p. 284), that at least partially, reflect a broad notion of ICC. With identified conceptual models, many have turned to operationalizing the construct so that it can be used in assessment, evaluation and research. The process of operationalizing a theoretical model typically involves transforming the primary sub-constructs into a series of individual survey items. Having multiple items to measure each sub-construct, which are part of the larger intercultural competence construct, helps minimize construct error (Groves et al., 2009). This error exists because intercultural competence is a latent, non-observable construct and relying on a single sub-construct or worse yet, a single survey item; leaves considerable measurement error as the construct is not fully specified.

The next section reviews five of the most commonly used survey instruments in intercultural competence research, specifically focussing on ones that have been used in study abroad research. Each instrument is introduced independently and includes a description of the theoretical underpinnings of the instrument, the key constructs or factors it utilizes, and some sample survey items. These are the construct and content validity approaches that researchers can make to show the strength of their instrument.

Intercultural Development Inventory. The Intercultural Development Inventory (IDI) is likely the most commonly used instrument in measuring the ICC of study abroad participants. The original version of the IDI was developed by Bennett and Hammer in 2002 and is based on the DMIS theory discussed in the previous section. The IDI

describes intercultural development as "attaining the ability to construe (and thus experience) cultural difference in more complex ways (Hammer, Bennett & Wiseman, 2003, p. 423). The IDI is comprised of approximately 50 items that respondents complete on a five-point scale, where 1 = strongly disagree and 5 = strongly agree. Two examples of sample items are “When I come in contact with people from a different culture, I find I change my behavior to adapt to theirs,” and “People from other cultures are not as open-minded as people from my own culture” (Paige et al., 2003, p. 472). Though the constructs in the survey are based on the six stages of intercultural development, the survey clearly focuses on primary elements of competence in detailing items that focus on attitudes, knowledge and skills, and behavioural aspects related to intercultural competence. The survey was developed by conducting interviews with more than 40 individuals asking them to make meaning out of their intercultural experiences. With that data, 200 intercultural statements were reviewed by four independent raters (Kappa 0.66) and developed into the six stages (Paige et al., 2003). A further seven intercultural experts reviewed these items and any items with an interrater reliability above 0.6 were retained. Factor analysis resulted in six factors all with reliability above 0.8. Additional validity work included some concurrent work, which found that the ethnorelative scales correlated positively with the World-mindedness Scale (Sampson & Smith, 1957) and the Intercultural Anxiety Scale (Hammer & Bennett, 1998; Stephan & Stephan, 1985). Paige et al., (2003) set out to conduct additional validity work in surveying 353 participants. Based on data from these participants, the original six factors loaded in the same way the original authors presented them, with Cronbach alpha values ranging from 0.74 to 0.91. However, Paige et al. (2003) found stronger statistical support for a basic two factor

model which would divide the ethnocentric and ethnorelative items appropriately. Predictive validity was found by testing participant variables like previous experience abroad and previous language experiences. In both cases individuals with more experience abroad or previous language training scored higher on the IDI (Paige et al., 2003). Analyses of the responses were not found to have any relationship with social desirability measures, suggesting that the IDI items are written in a manner that does not encourage socially desirable responses. The IDI is a commercial product, with a full training institute, summer courses, other related intercultural surveys and products for sale. To use the IDI survey an individual must first complete an intensive, multi-day qualifying seminar to help them become more aware of the properties of the IDI, including its reliability and validity. The IDI institute notes that they train individuals on how to interpret individual and group IDI profiles (Intercultural Development Inventory, 2017).

Global Perspectives Inventory. The Global Perspectives Inventory was developed by Braskamp, Braskamp and Merrill (2008). It defines a global perspective as the "capacity for a person to think with complexity taking into account multiple perspectives, to form a unique sense of self that is value based and authentic, and to relate to others with respect and openness especially with those who are not like her" (Braskamp, Braskamp & Engberg, 2013, p. 3). The GPI is based on two theoretical models. The first is Kegan's (1994) work on meaning making wherein individuals in trying to make sense of their lives, rely on their thinking, feeling and relating with others; or as Kegan put it, the cognitive, intrapersonal, and interpersonal. The second is Model of Intercultural Maturity described earlier by Baxter Magolda and King (2005). They

reframed Kegan's work into constructs in a student's social-cultural development and called the overall process the development towards intercultural maturity.

The GPI has three dimensions (cognitive, intrapersonal and interpersonal development) and each of the three dimensions has two sub-scales: knowing and knowledge, identity and affect, and lastly social interaction and social responsibility (Braskamp et al., 2008). Currently, the GPI is in its ninth revision. The original version was developed with pilot testing of 69 survey items that were given to nearly 130 participants. From their responses, factor analysis was used to reduce the number of items to 46. Using survey data from nearly 10,000 responses, additional factor analysis was conducted, and a six-factor model emerged with the factor loadings ranging from 0.66 to 0.78, which is far lower than the factors in the IDI instrument. Examples of survey items in this survey include the following from the Cognitive dimension "I consider different cultural perspectives when evaluating global problems," and from the Intrapersonal dimension "I enjoy when my friends from other cultures teach me about our cultural differences" (Braskamp, Braskamp & Merrill, 2013, p. 12). The rating scale used by the GPI is a five-point scale that ranges from 1 = strongly disagree to 5 = strongly agree and the scales range from having five to eight items.

In terms of developing reliability and validity arguments, the authors present several data points. The first is the test-retest reliability, which is the consistency with which individuals respond similarly when taking the survey again (Vogt, 2007). GPI test-retest correlations ranged from .59 to .73. An additional study by Anderson and Lawton (2011) did a comparative study of the IDI and GPI and found only small correlations ($r <$ .2) between the IDI and the GPI suggesting weak concurrent validity. However, that

study compared the individual factors and scales, not the overall scores of the respondents. The GPI has been used in numerous study abroad research studies (Engberg & Jourian, 2015; Gaia, 2015; Luchesi 2014) and is also used at colleges and universities with the entire student population to help measure how they develop interculturally during their time at university (Augustana College, 2012). This may be because the GPI, while still a commercial product, has a much lower price point than the IDI.

Global Mindedness Scale. The third instrument is the Global Mindedness Scale (GMS), developed as part of a dissertation (Hett, 1991). Global-mindedness is "a world view in which one sees oneself as connected to the world community and feels a sense of responsibility for its members. This commitment is reflected in attitudes, beliefs and behaviors" (Hett, 1991, p. 143). The instrument was developed using a triangulation process and is theoretically grounded in social theory construction (Schrag, 1967). To develop survey items, Hett conducted a broad literature review of related constructs and related empirical instruments and consulted with an expert panel of reviewers to help establish content validity once an initial set of items were developed (Hett, 1991). These reviewers had a content validity index of 0.88, suggesting broad agreement among them as relating to the primary construct named global mindedness (Golay, 2006). Following this, the survey was administered to a group of college students and then the psychometric properties of the scale were tested, resulting in a 30-item version.

The original version contained five dimensions of global mindedness. They are: Responsibility, which is a deep personal concern for people in all parts of the world; Cultural Pluralism, which is an appreciation of the diversity of the world's cultures; Efficacy, or a belief that an individual's actions can make a difference; Global Centrism,

which is thinking in terms of what is good for the whole world; and Interconnectedness, which is an awareness and appreciation of the interrelatedness of people around the world (Hett, 1991). The internal reliability for the GMS, using Cronbach's coefficient alpha was .90 overall (Golay, 2006). Alpha subscales ranged from .70-.79 (Hett, 1991 p. 102). The final version was a 30-item Likert–type scale, ranging across five choices from strongly disagree to strongly agree, like the IDI and GPI. One sample item from the cultural pluralism scale states, "The United States is enriched by the fact that it is comprised of many people from different cultures and countries." Another item from the global centrism scale states, "I sometimes feel irritated with people from other countries because they don't understand how we do things here" (Golay, 2006, p. 75).

Additional reliability and validity work available for the instrument include content validity work done by a set of four content experts. They had a content validity index (proportion of items determined to be related to the primary construct) of 0.88 (Hett, 1991). More recent validity work on the instrument highlighted that among teacher candidates there was a positive relationship between the GMS and a measure of teacher multicultural awareness (Asoclate, 2010). However, while Hett found that both gender and second language speaking ability correlated with GMS scores positively, Cui (2012) found no relationship among the variables.

Cross-Cultural Adaptability Inventory. The CCAI was developed to "assess your ability to adapt to living in another culture and to interact effectively with people of other cultures" (Kelly & Meyers, 1995, p. 1). It has been used in numerous study abroad research studies (Kitsantas & Meyers, 2001; Kitsantas, 2004; Mapp, 2012; Teranishi et al., 2008). The instrument was developed in a process like that of the GMS. The authors

conducted a broad review of the literature and created a checklist of characteristics found in the literature related to adaptation to and in other cultures. The researchers then had a panel of experts rate the importance of each characteristic. The resulting items were grouped into four scales: Emotional Resilience, Flexibility and Openness, Perceptual Acuity, and Personal Autonomy. The survey was initially tested and normed with a sample of 650 individuals who varied in age, education and occupation. 63% of the participants were male and 80% were U.S. citizens (Davis & Finney, 2006). The authors also share some additional reliability and validity work regarding the CCAI. Subscale reliabilities ranged from 0.68 to 0.90 indicating moderate to high levels of internal consistency (Kelly & Meyers, 1995; Vogt, 2007). Additional psychometric work in using principal components analysis was conducted by Davis and Finney (2006) who found poor support for the four-factor model presented by the authors. Moreover, they found that the factors in the instrument may have some discriminant validity issues in that there were substantial correlations between the factors. Additional work by Nguyen, Biderman, & McNary (2010) found stronger support for the validity of the instrument in the form of discriminant validity versus measures of the Big 5 personality dimensions, but they cautioned it might not be acceptable as a stand-alone instrument for expatriate staff selection, but better for training and awareness programs.

Summary. These are just four of the dozens if not hundreds of instruments that exist to measure and assess intercultural development (see Fantini, 2006 for a larger list). These instruments have numerous similarities. Overall, they tend to focus on knowledge, awareness and behaviours of individuals in intercultural situations. Though named differently, and measured with different sets of items and subscales, they are used in

study abroad to measure the notion of how individuals relate and communicate while abroad. This is not an argument that the instruments referred to above are not qualitatively and perhaps substantively different. However, in using these instruments, researchers are attempting to measure the impact of time abroad and identify if there are significant changes in student's intercultural attitudes, behaviour, skills and perceptions.

Other scholars in reviewing instruments used for intercultural research have come to similar conclusions. Fantini (2006) noted that at least 20 different terms were used for assessing ICC, and Sinicrope et al. (2007) reviewed many of these instruments and approaches highlighting their constructs and providing examples of sample items "paying particular attention to existing approaches and tools for its assessment" (p. 1). Spitzberg and Changnon (2009), who reviewed more than 300 theoretical models, theories and measurements relating to ICC, noted that there was "considerable similarity in their broad brushstrokes (motivation, knowledge, skills, context) and yet considerable diversity at the level of specific conceptual subcomponents" (p 35). This is exemplified in an eight-page table of the dozens of sub-constructs and factors related to motivation, knowledge, skills, and attitudes composure. Acknowledging the variability in specific items, it has been argued that any comprehensive model of intercultural competence would include motivation, knowledge, skills, context and outcomes (Spitzberg & Cupach, 1984; Spitzberg & Changnon, 2009). Models and instruments that include more of these would be a marker of quality. Overall, it is argued that there is extensive communality across ICC models and that "conceptual wheels are being reinvented at the expense of legitimate progress" (Spitzberg & Changnon, 2009, p. 45).

While there are numerous approaches and instruments to assess intercultural competence, they ultimately appear to measure a similar overall construct. Some models may focus more on attitudes, or knowledge or behaviours, and others are more comprehensive, but in their totality, they attempt to measure the knowledge, skills and attitudes needed by individuals to succeed in intercultural situations. As the previous sections have focused on conceptual and theoretical models of ICC and some of the more commonly used indirect or survey instruments, the next section reviews the empirical findings of researchers who have examined the relationship between the two constructs of interest: study abroad participation and intercultural competence.

Research on Study Abroad and Intercultural Development

It has long been suggested that study abroad is an effective means for improving intercultural competence (Hullihen, 1928; Hoffa & DePaul, 2010). The intercultural argument used to be one of several arguments in support of study abroad, which also included a foreign language and a broader academic and curricular argument. However, given the steady decline in the proportion of U.S. students who take courses in a second language, and the increasing movement towards short-term study abroad programs, the intercultural argument is a primary rationale advanced by a variety of stakeholders (Hoffa & DePaul, 2010; IIE, 2017). Whether for enhancing personal or national economic competitiveness, personal development or to aid in civic responsibilities, intercultural competence is an outcome that most stakeholders have become attached to and advance as a rationale for participating in study abroad (APCIES, 2012).

Given the importance that so many, and such a broad group of, invested parties ascribe to the value of intercultural competence, it is not surprising to see a

correspondingly large and growing body and community of research on the topic. The next section proceeds in four parts. First is a cursory review of the broad literature that looks at the links between study abroad participation and intercultural development. The second section examines, individually each of five student characteristics, including gender, program of study, year of study, nationality/ethnicity, and previous experience abroad. Following that will be a review of literature that includes the program characteristics identified by Engle & Engle (2003): length of student sojourn, entry language used in course work, context of academic work, types of student housing, provisions for guided/structured cultural interaction and experiential learning and finally guided reflection on cultural experience. These two sections form the basis of Astin's IEO model where inputs and environmental factors are considered in looking at outcomes. Finally, the section concludes with a focus on research design characteristics including the survey instrument, the research design (pre-post/pre-post with comparison group), study format (dissertation or article), the position of principal research (whether they were involved in the study abroad program), and finally study type (only quantitative or mixed methods study).

Regardless of the terminology used in reviewing study abroad research about intercultural development, it is impossible to ignore the fact that the research overwhelming finds that study abroad enhances intercultural competence. As described by Salisbury (2011), these studies may be grouped into three categories: studies that examine how study abroad impacts a) views of host culture or country; b) global perspectives or world-mindedness; and c) intercultural awareness or sensitivity of students. Studying abroad is purported to help students better understand the nuance of

the cultures in their home country supporting the contact hypothesis that engagement can help to reduce prejudice or increase understanding and empathy (Bicknese, 1974; Carlson & Widaman, 1988; Cushner & Karim, 2004; Nash, 1976). Bicknese used a simple pre-post survey model, while both Carlson and Widaman and Nash utilized comparison groups of students who did not study abroad. These additional measures strengthen a study against rival hypotheses, but still do not allow claims of causality.

A second category of studies looks at how study abroad participation is associated with changes in global perspective or world mindedness. As with views of host culture and countries, several studies have found that after study abroad, students have improved, or broader world mindedness (Chieffo & Griffiths, 2004; Clarke, Flaherty Wright & McMillen, 2009; Cushner & Mahon, 2002; Douglas & Jones-Rikker, 2001; Golay, 2006). The Chieffo & Griffiths study surveyed students only at the end of the study, relying on their estimate of how much they may have changed their worldviews rather than asking them to report their current worldviews. Clarke et al., while surveying students abroad and at home, used only post-tests. While the groups abroad did have higher scores on the measured attributes than the at home group at the end of the program, this does not negate the very real possibility that those who went abroad already had higher worldviews. The Clarke et al., study is a strong example of potential selection effects. That those interested in study abroad have a greater interest in world issues is in fact a finding that emerged in Carlson and Widaman (1998), who found students at home differed significantly from those who studied abroad even before their study abroad program. A threat to internal validity in many of these studies is a lack of a comparison group. This makes it

impossible to argue that studying abroad improves intercultural competence, as the students going abroad may well differ in meaningful ways from students who do not.

The third category of studies looks at how participating in study abroad impacts intercultural awareness and sensitivity. As with the two previous categories, numerous studies have found that studying abroad is associated with greater or improved awareness and sensitivity (Anderson, Lawton, Rexeisen, & Hubbard, 2006; Black & Duhon, 2006; Pedersen, 2009; Shaheen, 2004; Williams, 2005).

The above provides a brief introduction to the range of studies that exist on study abroad and intercultural competence. However, the relationship between the two constructs of interest, study abroad participation and intercultural development, is not nearly as direct or simple as many of the examples above suggest. The next section investigates research about study abroad and intercultural development with a specific focus on the characteristics of the study abroad students and how they moderate outcomes. Despite language sometimes used in the original articles, the articles in this literature review do not approach causality; the researchers were not able to randomly assign students to study abroad or not study abroad. Thus, the findings that are discussed are simply associations from the results of studies that used *t*-tests, and *t*-tests that used difference scores. There are numerous over-interpretations of the data in study abroad research, and this review will focus less on what the authors conclusions than on what their data and methodological choices indicate can be inferred from their results.

Study Abroad and Intercultural Development – Student Characteristics

Astin's IEO model posits that consideration of the relevant input or student-level characteristics are important for the role they can play in affecting the eventual outputs

and for highlighting the characteristics of those who do or do not participate in an activity, program or event (Astin & Antonio, 2012). In this review, six characteristics are examined: gender, academic program of study, year of study, ethnocultural/racial identity, nationality, and previous experience abroad. These are reviewed in order and in the following pattern. The first are studies that did analysis or comparisons specifically on the student characteristic. This could be a study that did a pre-post test of just the males or just the female students in a study to see if there is significant positive change in intercultural competence for both groups separately. Secondly, studies that were focused on a limited sample of students, such as a study that only looked at students in commerce, or only third year students are reviewed. The final category is studies that provided some descriptive details about the mix. For example, in this case studies that had students from several programs of study or a mix of years of study are examined.

Gender. When this study was proposed, no references were made about students who might identify outside of a male/female gender binary. Of the studies that were included in the final study, no studies specifically referenced any students identifying outside the male or female binary. I have identified only a few studies related to study abroad at all that have identified transgender and non-binary students in their descriptive results (see Bryant & Soiria, 2015; Kronholz, & Osborn, 2016; Niehaus & Inkelas, 2016). None of these studies however, met the inclusion criteria. A request was sent to IIE, the U.S. group who compile the most comprehensive statistics on study abroad participation among U.S. college students, to see if they ask institutions to report on the number of trans* identifying and, or, non-binary students participating in study abroad, but no response has been received as of February 13, 2019.

Participation statistics in published studies, dissertations and technical reports almost always report students identifying as male and female. But gender is also an important variable to examine given the historical trends towards increasing female participation in study abroad. Much has been made of the 'gender gap' in study abroad participation, leading many to ask what might be done to encourage more males to go abroad (Bond et al., 2009; Gore, 2017; Hurst, 2017; Shirley, 2006).

Few studies have explicitly compared how male and female intercultural development varies. Among those that have, the results are mixed. Female students in the Georgetown consortium study, using the IDI, reported statistically significant growth in intercultural development while abroad, while males had no significant change at all (Vande Berg, Connor-Linton, & Paige, 2009). Salisbury (2011) and Engberg and Jourian (2015) using gender as a covariate in regression models found small, negative, but not statistically significant relationship for male students as it relates to intercultural competence. In addition, three other studies found no significant differences between male and female intercultural development while abroad (Burrow, 2010; Massey & Burrow, 2012; Rexeisen, Anderson, Lawton, & Hubbard, 2008).

There were no studies identified that had only males or only females in the sample, thus the next way to examine the existing studies is to look at studies by proportion of female students: high (over 75%), average (60-75%) or relatively low (under 60%). There are studies that have high, average and low percentages of females in their sample that also had statistically significant findings in terms of overall intercultural development (Anderson et al., 2006; Black & Duhon, 2005; Clarke, Flaherty, Wright, & McMillen, 2009; Golay, 2006; Kitsantas, 2004). At the same time, there were also some

cases in which both high and average proportion female samples had no significant increase in intercultural development (Ingraham & Peterson, 2004; Patterson, 2006). However, except for Vande Berg et al., (2009), no study provided an effect size that might help the reader judge the magnitude of the difference between males and females. Thus, based on this evidence, making any statistical conclusions about a specific relationship between gender and likelihood or magnitude of intercultural development while abroad is not possible. Moreover, none of these studies provide a theoretical rational for why they conducted this analysis, what relationship they hypothesize is present, and what theory might underpin this relationship.

Program of study. Following gender, program of study is the second student level variable to examine. The students who originally went abroad in the University of Delaware program were all social science and languages majors. In the ensuing 100 years, the profile of students who go abroad and the programs they study in has changed significantly. The most recent statistics from Open Doors indicate that STEM students are now the largest group at around 26%, followed by commerce and business programs 21% and social sciences now down to 17% from 23% in 2010 (IIE, 2017). Foreign language (now combined with International Studies) comprises 7% of all participants (IIE, 2017). Few studies have specifically compared outcomes by program of study. Many studies either focus on students in a single academic program, most commonly professional schools or the studies include a mix of students from different academic disciplines. The Georgetown consortium study found that study abroad participants in humanities, social sciences or languages each had a statistically significant increase from pre- to post- test, while the other majors did not or in some cases (not mentioned) regressed (Vande Berg et

al., 2009). Sutton and Rubin (2004) found that students in commerce, journalism and social science majors reported greater global interdependence scores than did education majors. They also found that business students scored the lowest in term of cultural relativism compared with the other three programs of study. These differences conflict with reports by Burrow (2010) of visiting study abroad students to Canada and Massey & Burrow (2012) of Canadian students going abroad, which did not find significant differences in intercultural competence scores between Commerce or Arts & Science students.

Unlike the previous section in which there were no studies that examined samples of only women or only men, there are numerous studies of students in a single program of study. Commerce and business programs are the most common of these groups, but there are some other examples that include foreign languages, and very specific social sciences. Numerous studies of commerce students found intercultural development gains after studying abroad (Anderson & Lawton, 2008; Black & Duhon, 2005; Clarke et al., 2009; Rexeisen et al., 2009). Only Anderson et al. (2006) reported no significant change in a sample of only commerce students from a study abroad program. Among these studies, only Anderson and Lawton used a comparison group, meaning that it is possible the growth in intercultural development abroad could also have occurred to a similar group of students at home. What Anderson and Lawton did not provide were the statistics to see if the change from the commerce students was greater than that of the at-home group. Doyle (2009), in a study of only foreign language students, found that during the study abroad experience, students became “more comfortable and competent especially in cross-cultural interactions where traditions, practices, and customs may be unfamiliar or

markedly different," (p. 150). However, from this final study it is not clear how many or what percentage of students improved interculturally.

Other research included students from a wide variety of academic programs in their research. The findings among these studies are mixed though several found an overall positive significant increase in their intercultural competence measures (Braskamp, Braskamp, & Merrill, 2009; Kehl, 2006; Kitsantas, 2004; Nam, 2011), while others (Patterson, 2006; Williams, 2005) found no significant changes among study abroad students. In both latter studies the samples in the studies were quite small ($n < 30$), thus not finding a statistically significant effect is not entirely surprising. Patterson (2006) in fact noted that the students in their study had a small positive effect size ($d = 0.30$).

These studies include a mix of primarily positive results from research studies and dissertations. However understanding the impact of program of study from these studies cannot be determined from a narrative approach, which programs of study may have the most beneficial impact, and more importantly what the size of the relationship was. Moreover, there is no theoretical rationale provided for why students in a certain program of study might have intercultural competence development that differed significantly from students in another program.

Year of study. While some institutions or study abroad programs may have requirements or stipulations that students must have completed a certain number of years of study or be in a specific academic year, many programs are open to students who apply regardless of their year of study. The question researchers may be hinting at is if there is an optimal time in a student's undergraduate program for them to study abroad.

Both Clarke et al. (2009) and Rexeisen et al. (2008) studied third year undergraduates and both found that the intercultural competence scores for these students increased significantly by the end of the program. However, Anderson et al. (2006) in a study of only fourth year students did not find a significant difference in intercultural competence between the start and the end of the program. In each case, the samples are quite small ($n < 25$). Most studies, especially ones with larger samples, include samples of students from a variety of academic years of study. Numerous studies with mixes of participants have reported significant changes in intercultural development (Anderson & Lawton, 2008; Kehl, 2006; Kitsantas, 2004; Nam, 2011; Salisbury, 2011; Vande Berg et al., 2009). Only a few studies (Burrow, 2010; Massey & Burrow, 2012) included variables related to year of study and found no significant differences from the pre to post test between the varying years of study of the participants. Given that many academic programs follow a sequential or scaffolding curriculum structure, using year of study makes more sense than using age of participants. Moreover, if students must acquire specific credits during their time abroad to maintain their ability to progress in their program of study, it likely makes more sense to focus on year of study, rather than the participants' age.

Ethno-cultural identity. Broadening study abroad participation to be more reflective of the general population of students in college and university has been a long-standing goal of many professional organizations and scholars (Desoff, 2006; Stroud, 2010; NAFSA, 2011; Woolf, 2007). The table below provides the most recent comparisons of study abroad and post-secondary education participation (National Center

for Education Statistics, 2016). No comparable data could be found for Canada or the European Union.

Table 2

Percentage of U.S. Study Abroad by Ethnicity

Race/Ethnicity	U.S. Post-secondary Enrollment 2015-2016	U.S. Students Abroad 2015-2016
African American or Black	14.1%	5.9%
Asian/Pacific Islander	6.8%	8.4%
White	57.6%	71.6%
Hispanic/Latinx American	17.3%	9.7%
Multiracial	3.5%	3.9
American Indian/Alaska Native	0.8%	0.5%
Non-resident/undocumented	3.5%	Not Available

Many studies reviewed (see exceptions Edwards, 2009; Jackson, 2008; Kehl, 2006; Nam, 2011) reported having a higher proportion of White students than the IIE reported, making them even less representative of the general population of higher education students. Many studies that have reported ethnicity or race in their demographic characteristics, regardless of the proportions, have also reported overall statistically significant increased in intercultural competence, (see Angulo, 2008; Golay, 2006; Ingraham & Peterson, 2004; Kitsantas, 2004;) but no studies found compared students by their ethnicity.

These studies however vary in the instruments used, research design (pre-post versus post-test only), use of control groups, study abroad locations and sample sizes. Thus, making any conclusions regarding how ethnocultural identity may be associated with intercultural development not possible. It is accurate to say that the ethnicity of

participants is not reflective of the population of higher education in general, and that the published studies might be even less representative at that.

International or domestic student participation. The next two characteristics begin to investigate the role that previous life experiences may have on the relationship between study abroad participation and intercultural development. The first of the two is whether the study abroad student is considered an international or a domestic student. Kehl (2006) noted that his sample was 98% local or domestic students, while Massey and Burrow (2012) had 92% domestic students in their sample. Neither found any significant differences between the changes reported by domestic and international students. Compared to many of the characteristics described above, fewer studies are reporting the proportion of students who are domestic or international studies, but it is also possible that there were, especially for older studies, few if any international students studying abroad.

Previous experience abroad. The next section examines the degree to which study abroad participants' previous international experience (work, study, travel) may be related to intercultural development. The question is not which students ended up with the highest intercultural scores or measures, but if previous experience abroad, or the lack of it, supports or inhibits intercultural development while studying abroad. This category includes data of students who worked abroad, travelled abroad and studied abroad; the underlying issue is if this previous experience is related to an individual's ability to develop interculturally.

The Georgetown Consortium study examined previous experience abroad in general and previous experience studying abroad. The authors found no significant

increase in intercultural competence among students with any previous international experience and specifically no effect for those who had previous study abroad experience compared to those who had no experience abroad or never studied abroad previously (Vande Berg et al., 2009). Other results are more mixed. Patterson's (2006) study noted that 38% of participants had previous international experience while Anderson et al., (2006) indicated that in their sample 43% had been abroad. Neither study reported an significant positive change in intercultural competence however and they did not separate the sample to test to see if those with and without experience abroad had significant differences. These findings contrast with three positive findings of intercultural development in studies that had a large percentage of participants (between 50% and 80%) reporting some previous experience abroad (Kehl, 2006; Nam, 2011; Rexeisen, Anderson, Lawton, & Hubbard, 2008). Interestingly, only Vande Berg et al. (d = 0.21) and Patterson (d = 0.30) provided an effect size to help understand the magnitude of the effects. Thus there are two examples of studies with a large percentage of students with previous experience abroad that reported statistically significant increases in intercultural competence. This finding may suggest an association between previous experience abroad and intercultural competence gains.

Summary: Student characteristics. The review above investigated studies that compared, focused exclusively on, or reported descriptive statistics of six different student level variables that could be related to the development of intercultural competence in a study abroad program. Throughout the reviews, a mix of significant and not significant findings were reported. More importantly, within these sections and studies, even greater variability in the range of these student variables was reported. The

six characteristics relate primarily to the I in Astin's IEO model, but previous experience abroad is also related to Allport's Intergroup Contact Theory, as students with previous intercultural experience may be more likely to benefit than those without.

While six variables were studied in the previous section, numerous other variables were not consistently featured in descriptions of students participating in a study abroad program. These include whether or not a student had an accessibility/disability, which might further support the notion that study abroad is primarily designed for students without accessibility challenges. While some studies provided details about the percentage of students who were international, very few included information about how many students identified as first in their family to attend college or university. A third category consistently omitted was socio-economic status. While some studies (see Salisbury, 2011, Stroud, 2010) specifically include variables relating to receiving student aid or parental income, the majority did not. These are just three omissions of student characteristics, which may not necessarily relate to the development of intercultural competence, but do speak to the profile of who is participating in study abroad or who is interested in study abroad. They are also three variables that relate strongly to the justified equity, diversity and inclusion concerns many have about who participates in study abroad. A second set of variables relate to overall intent to study abroad (Salisbury, 2011), motivation or goals to study abroad (see Burrow, 2010; Kitsantas, 2004) or what Engberg and Jourian (2015) have named intercultural wonderment, the curiosity or interest in seeking out new experiences while abroad. These variables may help better understand which students are more likely, more motivated or more interested in developing interculturally while abroad.

Attempting to summarize the student characteristics findings, using only a descriptive approach, where the authors describe the results of a series of related articles, makes it challenging to provide defensible conclusions about how the relationship between study abroad and intercultural competence development differs by student characteristics. In addition, this review primarily examined the variables independently, ignoring the possible correlations between variables like gender and program of study. Though most studies appear to have significant results or indicate that during the study abroad program students did increase or improve their intercultural competence, very few provided effect sizes to help readers understand the magnitude of these changes. For that to occur, the results must be reviewed statistically, not narratively.

Study Abroad and Intercultural Development–Program Characteristics

The previous section focused on the student level characteristics, or the inputs in Astin's IEO model. This section of the literature review focuses on the study abroad program characteristics that differentiate one program from another. I examine six of the seven characteristics identified by Engle and Engle (2003). The only variable not studied is entry language target competence as only one study (Vande Berg et al., 2009) identified it.

Duration. Of all the variables examined in study abroad literature, the role of duration is likely the one that has received the most attention among published studies, dissertations and reports. Recall that the initial junior year abroad programs at the University of Delaware were for an entire academic year. Since these earliest days of formal organized study abroad, the duration that students have been going abroad

continues to shrink. In 2015-16, 62% of U.S students were abroad for 8 weeks or less, 34 % between 8 weeks and 16 weeks and just 4% for two semesters or more (IIE, 2016).

Several studies have specifically compared how intercultural competence changes for different durations of study abroad programs. Some studies have found that longer duration abroad was associated with greater intercultural development (Dwyer, 2004; Ingraham & Peterson, 2004). Other studies are more nuanced in saying that programs that were approximately a semester in length, compared to full year and short-term programs had the largest intercultural competence gains (Vande Berg, et al., 2009), while others found that both eight weeks and full-term study abroad programs had the same intercultural development (Kehl, 2006). By contrast, Burrow (2010) and Massey and Burrow (2012) found no significant differences in intercultural competence development for students who studied abroad for one term, or for student who studied two terms.

Other studies have focused on a program or programs of similar length. One set of findings with students in programs less than eight weeks in length did not report any significant increase in intercultural competence (Black & Duhon, 2006; Braskamp et al., 2009; Kurt, Olitsky, & Geis, 2011; Nam, 2011). However, this was not always the case and some research of short-term study abroad programs did not find any significant increase in intercultural competence (Anderson et al., 2006; Patterson, 2006). Some studies of programs that were one semester in length reported that participants' intercultural competence significantly increased during their sojourn abroad (Anderson & Lawton, 2008; Golay, 2006; Rexeisen, Anderson, Lawton, & Hubbard, 2008), though this finding is not uniform in the literature, as other studies of programs that were one

semester in length did not report any significant change (Chieffo & Griffiths, 2004; Engle & Engle, 2003; Medina-Lopez-Portillo, 2004).

Whether students who participate in short-term study abroad programs can significantly increase intercultural competence and if longer study abroad programs lead to even greater growth is a topic that remains uncertain. The above highlights that while most studies, regardless of length, are associated with improvement in intercultural competence, this is not uniform, and there are many cases where the longer programs have no effect while some shorter programs did. This analysis, of course, ignores the many ways in which these longer and shorter duration programs, and the students who participate in them, may vary.

Language used in course work. The second variable examined is the language used in coursework while abroad. In this variable, I am looking at whether a program was taught entirely in English, primarily English with introduction to the host language or entirely in the local language. Recall that less than 10% of students indicated that they were foreign language or international studies majors. Among studies where some students studied in a foreign language and some in English, Vande Berg et al. (2009) found students who were studying in a foreign language gained significantly more on the IDI than those who did not. By contrast Spenader & Retka (2015) had eight separate groups of students studying abroad. Five of those groups studied in another language, three in English only. Of the five in another language, two had significant gains and three did not; of those studying in English, one of three groups improved significantly in their IDI scores. In studies only of those in another language, some found significant gains including Edwards (2009) in study of Japanese students studying in English, as did Smith

& Moreno-Lopez (2012). But two samples of students studying exclusively in Spanish by Medina-Lopez (2004) did not result in a significant increase in intercultural competence between the pre-test and post-test. There were numerous studies of students studying abroad in their primary language that did have significant results including Anderson and Lawton (2011), Armfield (2004) and Kitsantas (2001). Studies that had a mix of those in a foreign language and studying in their primary language did not always have a significant overall change in intercultural competence from the pre-test to the post-test (see Burrow, 2010; Hoff, 2010 for examples). The relationship between intercultural competence development and language of instruction, like many of the other variables appears to be mixed and clear conclusions about how it relates to intercultural competence development is uncertain.

Faculty instruction. After examining the language used in coursework, the next section examines the instruction. This focuses on who is teaching the courses abroad; faculty from the home country, the host country or a combination of the two. Only a single study could be identified that specifically compared students in these different types of academic programs. That study found that those who took classes with home faculty or a combination of host faculty and home faculty reported greater intercultural competence growth than did those who studied primarily under host faculty (Vande Berg et al., 2009). This finding suggests that students abroad develop more interculturally when their program is led by faculty more, not less, familiar with their own culture. This would appear to conflict with Allport's theory, but the study also includes considerable out of class intercultural contact, which does support the theory.

Most other published studies that provide details of the faculty teaching the course or program abroad indicated that the program was led by faculty from the home country. In most of these cases the studies reported some intercultural growth during the study abroad program (Black & Duhon, 2006; Clarke et al., 2009; Nam, 2011). These findings are supported by two other studies, which found that study abroad students who were taught exclusively by host faculty reported no significant changes while abroad (Burrow, 2010; Massey & Burrow, 2012). The third category looks at research studies that had a mix of host and local faculty. In these cases, most research seemed to indicate that these arrangements had positive impacts on intercultural development while abroad (Kitsantas, 2004; Rexeisen, Anderson, Lawton, & Hubbard, 2008).

Housing. The fourth category looks outside of the classroom towards the housing arrangements of the students. This is potentially an element of intercultural contact, as study abroad students might live in a homestay with a local family, share an apartment with local or international students, or they may just live on their own or with others from their home country. Like academic context, student housing is a partial proxy for intercultural interaction.

One of the most important studies looking at housing and intercultural development was done by Vande Berg and colleagues (2009). They found that living with other American students had a significant impact on intercultural development and small effect size ($d = 0.32$). By contrast, living with a host family or host nationals was not significant. Most studies, which have reported on housing options, tend to have all students in similar accommodations as they are studies of individual study abroad programs. Students in the research conducted by Anderson et al. (2006), Anderson and

Lawton (2008) and Rexeisen et al. (2008) were all placed with homestay families; however only in the two latter cases was the study abroad program associated with a significant increase in intercultural competence.

Cultural interaction and experiential education. The fifth category looks at the degree to which the students' study abroad program provided specific faculty or administrator led cultural interaction, guided tours and/or experiential learning opportunities while abroad. The studies, which provided details of these activities, varied significantly in the range and type of experiential and intercultural involvement that the study abroad students had. For example, Clarke et al. (2009) reported that their students "visited businesses, governmental institutions, and cultural sites in six Western European countries" (p. 176). Kitsantas (2006) noted that the programs she studied had a mix of guided and unguided off-campus activities, wherein "local and more distant excursions were provided for the students and most study abroad programs allowed a three-day weekend for exploration" (p. 444). Finally, Anderson et al. (2006) described that the study abroad program in England

> [I]ncluded guest speakers, company site visits, and travel to local cultural sites. The program included a series of lectures by a British professor entitled ''British Life and Culture'' that covered topics such as British politics and Parliament, the National Health Service, taxation, and an overview of the EU. The program also included site visits to Cambridge and Canterbury while in England (p. 462).

Despite the numerous guided activities provided in the Anderson et al. study, the two other studies were those in which students showed significant improvement in intercultural competence. Understanding the nature of campus interactions and the quality of the interactions of students is challenging unless extensive descriptions are provided as

well as additional survey information on the frequency or intensity of some of these outings.

Guided reflection on cultural experience. The sixth category extends the previous one. While off campus activities are believed to increase intercultural contact and interactions, some have argued that it is the reflection of those interactions that enhances intercultural development (Vande Berg, 2009). The Vande Berg et al. (2009) study measured reflection by the amount of time students reported spending with their cultural mentor while abroad (face to face or virtually). Their research highlights that as students spend more time interacting in guided reflective activities, their intercultural competence scores increase compared to when they began the program (d = 0.35). Other studies did not focus specifically on the amount of guided reflection but indicate that part of their study abroad program involved a service-learning component. These studies which indicated that the study abroad program included service-learning elements, of which reflection is a key component, all reported significant growth in intercultural development (Anderson & Lawton, 2008; Nam, 2011; Rexeisen, Anderson, Lawton, & Hubbard, 2008). Though few studies have explicitly focused on guided reflection on their experience abroad, the few that have seem to report significant overall results.

Summary: Student abroad program characteristics. The review above focused on six different characteristics of study abroad programs. Variables in this section largely follow the structure from Engle and Engle's study abroad program classification framework. They are the E (Environment) in Astin' IEO model and many (especially housing, faculty, nature of study abroad program, cultural interaction and duration) are related to Allport's Intergroup Contact theory. The program characteristics are all ways a

student can be closer to learning about the local culture, customs and where possible language; all of which Engle and Engle likely would argue would enhance the opportunity to develop intercultural competence. The Engle and Engle model, however, does not include specific variables for pre-departure orientation, on-site orientation and ongoing support but instead combines most of these activities into the final category on guided reflection on cultural experiences. They are also less clear about the role of faculty and staff interactions with students. It appears they see a place for staff and faculty in supporting reflection and interaction, but their description does not entail how or what role they have in supporting students while abroad. For example a student who potentially has visa or immigration issues once in the country could find the situation overwhelming and their ability to focus on intercultural competence would be jeopardized.

Just as with the student characteristics variables, the ability to investigate how these program characteristics may moderate the relationship between study abroad and intercultural competence depends on the willingness of researchers to provide descriptive and narrative details about them. Like with the student characteristics, making confident generalizations about the study abroad program characteristics that are associated (positively or negatively) with changes in intercultural competence cannot confidently be done using only a narrative approach.

Study Abroad and Intercultural Development – Research Design Characteristics

The third set of criteria for analysis involves choices the researchers make in conducting their study. There are five criteria studied here. I review them more briefly than student or research design characteristics, as they have not been the specific focus of

published research. These decisions by themselves do not impact a student's ability to develop interculturally, however the decisions a researcher makes can have a meaningful impact on their overall findings and conclusions.

Publication type. This category focuses on the nature of the publication whether as a dissertation or masters' thesis, journal article or referred paper from a conference. The question being tested here is if articles and referred conference papers have different, possibly smaller effect size due to additional faculty research experience and possibly more rigorous review process than thesis and dissertations.

Researcher involvement. The second characteristic attempts to identify if the researcher was also a participant in the program, typically as a faculty member, or in some cases as an on-site mentor or advisor. This variable is included to see if studies where the researchers was at least arm-length from the study abroad program itself have more conservative findings than ones where the researcher was part of the program.

Survey instrument. This looks at which of the standardized instruments reviewed earlier the study used (Intercultural Development Inventory, Global Perspectives Inventory, Global Mindedness Scale, Cross-Cultural Adaptability Inventory) or if they used another instrument or created and developed their own. Lengthy descriptions of some of the most common survey instruments is presented in a previous section on operationalizing intercultural competence. This study will include a variable on survey instrument use as it is believed that these could be related to the magnitude of change from pre-test to post-test. A handful of studies have used more than one survey instrument on the same population. Anderson et al. (2011) had a population of students (n = 39) complete both the IDI and the GPI in a pre-test and post-test and Williams in their

2002 dissertation had 52 (27 abroad and 25 at home) students complete both the Intercultural Sensitivity Index and the Cross-Cultural Adaptability Inventory. Anderson found statistically significant increases on both instruments, as did Williams. Converting their results to effect sizes in the Anderson study the GPI produced the larger effect size and for Williams the ISI was larger. These two examples highlight how the same population with same study experience abroad can produce different estimates depending on the instrument they chose. However, this approach may be prone to testing effects and I would be curious if the researchers varied which instrument was given first to the students in each case to mitigate these potential effects.

Research design. The last characteristic focuses on the groups and the timing of the surveys. There are three options. The first is the traditional pre-post with students surveyed at the beginning and the end of the study abroad program. The second expands the first option by adding a comparison group who are also surveyed at the start and end of the same time period. A final type is a retrospective pre-test, wherein the authors conduct the pre-test at the same time as the post-test and ask participants to reflect on their experiences and ascertain how they would have rated themselves at the start of the program. Studies that use measured intercultural competence of students after a study abroad program, even when they included a comparison group, are not included in this dissertation. Of the research design characteristics above, only the research instrument has ever been the focus of a study before. This is because research studies tend to only have a single research design or publication type.

The variables in this section relate to the validity framework and specifically threats to internal validity (maturation and selection) and instrumentation concerns. They can impact the overall relationship which is guided by Allport's theory

Conclusion

Determining which characteristics have the most impact on the development of intercultural competence is challenging. Simply tallying the number of studies with significant or non-significant findings could bring more clarity around the issues than a narrative approach. This vote-counting approach would have researchers rely on counts of the studies reviewed and conclusions are reached based on the number of votes or studies that fall into the significant, positive, negative or non-significant piles. With this information, the researcher determines the veracity of previous claims, and their direction, by how many studies occupy each pile. The problems with this approach are numerous, most notably a reliance of significance and the blunt way of saying that something has a positive, negative or no effect, rather than examine the magnitude of the effect (Hunt, 1997; Card, 2012). This dissertation goes beyond the vote-tallying approach by assigning weights to each study based on their precision and by calculating the magnitude of the study conclusions so that they can be analyzed quantitatively.

Hypotheses

As a result of the preceding literature review and the guiding theoretical framework, the following hypotheses are proposed from my three research questions.

Research question 1. Does an overall relationship exist between study abroad participation and intercultural competence? Is there is there significant variation in this relationship across studies.

The hypothesis is that the study will find a significant, positive, small effect size (($0.20 < g < 0.50$) regarding the relationship between study abroad and intercultural competence and that there will be significant variation or heterogeneity among effect sizes.

Research question 2. To what extend do individual student, study abroad program and research design characteristics moderate the relationship between study abroad and intercultural competence. Table three provides an overview of the hypotheses by each set of characteristics.

***Student characteristics*.** It is hypothesized that no student level characteristics (gender proportion, program of study, year of study, ethnocultural identity, international/domestic status or previous experience abroad) will moderate the mean study effect size.

***Study abroad program characteristics*.** It is hypothesized that as the characteristic move towards Engle & Engle's level five we would see larger effect sizes except duration. That is we expect to see large effect sizes associated with, programs in a foreign language, students taught by host faculty, students in direct enrolment rather than in collective groups from the home institution, students in homestays, students who experience intentional cultural interaction and whose programs have guided reflection.

***Research design characteristics*.** It is hypothesized that studies in dissertations, studies where the researcher was a participant in the study abroad program would both be associated with larger effect sizes. In considering instruments used to measure intercultural competence, I hypothesize that studies using the Intercultural Development Inventory will have lower effect sizes than other instruments as it is the most tested

survey tool. Finally it is expected that studies that use a retrospective pre-test would have the largest effect sizes, followed by studies with a traditional pre-post design and the smallest effect sizes for studies that used a pre-post comparison group design.

Table 3

Research Question Two Hypotheses

Variable	Hypothesis
Student Characteristics	
Gender	No relationship with mean effect size
Program of study	No relationship with mean effect size
Year of study	No relationship with mean effect size
Ethnocultural identity	No relationship with mean effect size
International/domestic	No relationship with mean effect size
Previous experience abroad	No relationship with mean effect size
Study Abroad Program Characteristics	
Duration	No relationship with mean effect size
Language of instruction	Foreign language instruction positive association
Faculty instruction	Taught by host faculty positive association
Nature of study abroad group	Direct enrollment/small groups positive association
Housing	Homestay positive association
Cultural interaction	Cultural interaction positive association
Guided reflection	Guided reflection instruction positive association
Research Design Characteristics	
Publication type	Articles positive association
Researcher involvement	Researcher involvement positive association
Survey instrument	Intercultural Development Inventory negative association
Research design	Retrospective pre-test, followed by pre-test/post-test design, followed by pre-test/post-test

Research question 3. To what degree do combinations of student, study abroad program and research design characteristics, and all characteristics explain variance in the mean study effect size? Table four provides a summary of the hypotheses for all four planned analyses and the expected findings.

This analysis involves four sets of meta-regressions. The first is that in a meta-regression model student characteristic will not moderate the mean study effect size. The second question relating to program characteristics, I predict that cultural interaction and guided reflection will be significant positive moderators, and that duration will have no relationship. The third meta-regression I hypothesize no relationship for publication type or researcher involvement but a negative relationship for studies using the IDI and larger effect sizes for studies using a retrospective pre-test. A final combined meta-regression will be run using variables from all three previously meta-regressions. It is predicted that study abroad program characteristics will have the most significant relationship followed by research design and no relationship for student characteristics.

Table 4

Research Question Three Hypotheses

Variable	Hypothesis
Student Characteristics	No overall relationship with mean effect size
Study Abroad Program Characteristics	
Duration	No relationship with mean effect size
Language of instruction	No relationship with mean effect size
Faculty instruction	No relationship with mean effect size
Nature of study abroad group	No relationship with mean effect size
Housing	No relationship with mean effect size
Cultural interaction	Cultural interaction positive association
Guided reflection	Guided reflection instruction positive association
Research Design Characteristics	
Publication type	No relationship with mean effect size
Researcher involvement	No relationship with mean effect size
Survey instrument	Intercultural Development Inventory negative association
Research design	Retrospective pre-test, followed by pre-test/post-test design, followed by pre-test/post-test
Combined Meta-Regression	
Student Characteristics	No relationship with mean effect size
Study Abroad Program	Positive relationship with mean effect size

Characteristics	
Research Design Characteristics	Small positive relationship with mean effect size

The next chapter outlines an approach to systematically reviewing the quantitative research relating to study abroad participation and intercultural development to understand the conditions associated with the variations in the findings.

Chapter 4: Methodology

Chapter four details the process of conducting a meta-analysis of the relationship between study abroad participation and intercultural development. This includes the method of obtaining the studies in this paper, the process used to determine their inclusion/exclusion, the criteria for inclusion, the coding structure and trustworthiness approach, the analytical approach to computing effect sizes and answering the following research questions:

1. Does an overall relationship exist between study abroad participation and intercultural competence? Is there significant variation in this relationship across studies?

2. To what extent do individual student, study abroad program and research design characteristics moderate the relationship between study abroad and intercultural competence?

3. To what degree do combinations of student, study abroad program and research design characteristics, and all characteristics explain variance in the mean study effect size?

Before describing the analytic process, I introduce meta-analysis with a brief review of the history, the logic, the rationales and weakness in using a meta-analytic approach.

An Introduction to Meta-Analysis

History of meta-analysis. Though the application of meta-analysis is still in its infancy in many research circles, the roots of this analytical technique date back more than 100 years (Hunt, 1997). The earliest references of its underpinnings as a methodological approach are credited to the British mathematician, Dr. Karl Pearson (Hunt 1997, Lipsey & Wilson, 2001). Pearson developed a technique for combining divergent results from inoculations against typhoid fever and observed that “the fact that the correlation, while sensible … is subject to remarkable variations” justified further

investigation (Simpson & Pearson, 1904, p. 1244). They noted that with small samples, the variance in the results might be attributable to chance, meaning that the true effectiveness of inoculation was not known. In the 1930s, researchers were experimenting with farming techniques, attempting to understand the conditions and approaches that would lead to the best yields (Hunt, 1997). Tippett (1931) collected data from each of the experiments, including sample size, difference in crop yield between farming techniques, and measured the amount of variance in yield that occurred by chance within a specific technique. With this information, he was able to control for sample size and then compare the difference between farming techniques to the difference within technique and discern how likely it was that the results from all the studies were by chance.

In the mid-70's Robert Rosenthal (social psychology) and Gene Glass (education), were independently developing techniques for the combination of studies (Glass, 1976; Robinson, 2004). The spring of 1976 was a seminal moment in the advancement of meta-analysis, when Glass, then president of the American Educational Research Association, planned to deliver his presidential address on his advancement of the topic (Robinson, 2004). At that address, in front of more 1,000 AERA members, Glass delivered a presentation on meta-analysis and the audience "was blown away by it. There was tremendous excitement about it; people were awestruck" (Robinson, 2004, p. 5). Glass described a five-step meta-analytic process, which still forms the foundation of meta-analytic research.

Rationale and logic of meta-analysis. Reviews of research can lead to a deeper understanding of the existing knowledge base, the approaches previous researchers used and the strengths and weakness of their approaches and applicability and limitations of

their findings. However, the ability of any individual researcher to be informed of all the research even in the narrowest of fields is questionable, especially at a time when new research studies, dissertations, conference presentations appear daily. In addition, as the volume of research increases in any given field, so do the methods, measures and samples researchers' use to conduct this research (Bowman, 2012). Thus, attempting to sift through these volumes of research to describe narratively the patterns and conclusions in them, not to mention understand what factors may account for the disparate findings, is increasingly difficult and challenging.

Description and definition of meta-analysis. Before proceeding further, a few additional descriptions are required to delineate what differentiates meta-analysis from primary analysis and from secondary analysis. Primary analysis refers to what is commonly thought of as original research, whereby researchers collect data from individuals using qualitative or quantitative approaches (Card, 2012). It also includes reviews of existing policy documents and websites. By comparison, secondary analysis refers to the 're-analysis' of data, often to answer different or new research questions while using existing data. Secondary analysis may be performed by the original research team, or by new researchers using the original findings. The use of original, or raw, data is characteristic of both primary and secondary analysis (Lipsey & Wilson, 2001).

Meta-analysis involves the statistical analysis of the results of existing empirical studies. This differs from primary or secondary analysis in two ways. First, meta-analysis uses previous studies and typically analyzes aggregate data, rather than raw data like in primary and secondary research. Meta-analysis utilizes the effect sizes of existing reports, which if not published in the original or secondary report, can usually be calculated using

other information in the study. Secondly, meta-analysis involves the combination of two or more empirical studies and typically draws inferences from samples, rather than from individuals. This combination of multiple studies is what leads to its comparisons to literature reviews (Borenstein, Hedges, Higgins & Rothstein, 2009; Card, 2012; Lipsey & Wilson, 2001).

Meta-analysis as literature review. The literature review is a feature of virtually any empirical article, dissertation or conference presentation. The review is an investigation or synthesis of existing literature related to the present study (Card, 2012). Effective literature reviews summarize the findings, strengths and weaknesses of the previous work, typically resulting in a presentation of a 'gap' or space that justifies a need for the planned study (Creswell, 2012). Literature reviews however differ in their focus, goals, coverage, depth, organization, and method of synthesis (Cooper & Hedges, 1994). It is the approach to synthesis that particularly distinguishes meta-analysis from three other forms of research synthesis and literature reviews.

The first and most common approach is the narrative research review. In this approach, the review is an evaluation of selected research publications related to their constructs of interest. Typically, the focus is on the findings, pointing to methodological strengths and weakness, the populations examined and concluding with a summary about what is known, not known and is still needed to be known about those chosen constructs. The exact process of how the researcher arrives as these conclusions is not often stated. It is rare that these forms of review provide any examples of the approaches to ensuring trustworthiness that they use in developing their review and conclusions. Instead it is left

to the reader to judge the researchers' ability to synthesize and make meaningful conclusions and is highly prone to subjectivity (Card, 2012; Hunt, 1997).

A second approach is informal vote counting. In this approach, the researcher relies on a count of the studies reviewed and conclusions are reached based on the number of votes or studies that fall into the significant, positive, negative or non-significant piles. From this information, the researcher determines the veracity of previous claims, and their direction by how many studies end up in each pile. The problems with this approach are numerous, most notably a reliance on only the *p*-values to arrive at a very blunt way of determining if a relationship has a positive, negative or no effect. Recall that a "*p*-value is defined as the probability, under the assumption of no effect or no difference (the null hypothesis), of obtaining a result equal to or more extreme than what was actually observed" (Goodman, 1999). In this approach, nothing can be said about the size of the effect, nor what conditions or variables may moderate its strength or even direction (Card, 2012; Hunt, 1997).

The third approach is also a vote-counting process, but one that uses statistical analysis of the *p*-values from published research. Formal vote counting relies on analysis to determine if statistically significant research results appear more than the expected 5% of time (using a traditional type I error rate of .05). The problems with this approach are like the previous example. Statistical analysis of *p*-values may be better than simply counting them, but it obscures what researchers should be concerned with; that being the direction and the magnitude of the relationship (Borenstein et al., 2009). Borenstein noted that power, the ability of a study to detect a statistically significant effect, in many fields of research is very low and that this has not changed over time. Non-significant *p*-values

(those where $p > .05$) may appear when studies are underpowered and have small samples, even though they may have an effect size that suggest a promising practice or approaching. The meta-analytic approach reoriented the focus on study outcomes from *p*-values to effect sizes and will be discussed in greater detail in the next section.

The basics of meta-analysis. While meta-analysis has evolved considerably since Gene Glass gave his presentation at AERA in 1976, the basic characteristics he introduced remain valid now more than 40 years later. First, meta-analysis focuses on empirical studies (Borenstein et al., 2009; Card, 2012). It cannot be used to analyze theoretical, conceptual, or policy reviews. Secondly, meta-analysis includes only research studies that result in quantitative outcomes and use quantitative methods. This can include descriptive or inferential statistics in summarizing and presenting data. It precludes other forms of research such as qualitative, case studies and ethnographic work (Card, 2012). It should be noted that meta-analysis does not discount the considerable value and contributions that these other forms of research add to the field of study, only that in the meta-analysis process they cannot be included. Findings from qualitative and ethnographic research play a vital role in determining study characteristics to code and coding schemes to employ. Thirdly, meta-analysis is an analytical approach that first involves the coding and analysis of statistics from previous studies to develop a data set. Fourthly, meta-analysis requires that studies of similar concepts, constructs or relationships be compared. Studies that include diverse concepts or relationships can be argued as not originating from the same population of studies. Still, regardless of the relationships investigated, what is common across all meta-analysis is a singular focus on the effect size. This underscores the notion that effectively all empirical studies produce a

result, but that they vary in direction (positive or negative) and in magnitude (very small to very large). Therefore, a meta-analytic approach recognizes the important contribution that studies with non-significant results (*p*-values above .05) make in understanding phenomena. In fact, these studies are as equally important to those studies with statistically significant results as meta-analysis intentionally focuses on all studies that meet the inclusion criteria regardless of their *p*-values (Lipsey & Wilson, 2001).

Strengths of the meta-analytic approach. The benefits and rationales for using a meta-analytic approach, over a single study or narrative review fall into four categories (Borenstein et al., 2009; Card, 2012; Hunt, 1997). Meta-analysis establishes a clear protocol and discipline in the reporting and summation of research findings. Meta-analysis involves a structured approach that requires extensive documentation that can be critiqued and/or replicated. This includes specification of inclusion criteria, which helps the reader understand what studies were included and why additional studies were excluded. A meta-analysis also provides definitions of each of the key variables of interest and full coding details and procedures for how the data set was developed.

A second benefit of meta-analysis is that, unlike conventional reviews of literature, it is a way to represent a body of literature in a differentiated and statistically defensible way (Card, 2012; Lipsey & Wilson, 2001). While narrative literature reviews present the literature through a description of the primary constructs and ideas of interest, and ideally a critical review of the existing literature, it is not often clear which criteria the researcher values and which they reject or support in presenting their analysis of the literature. The coding procedure of creating effect sizes shows both the magnitude and direction of the previous findings, whereas traditional vote counting may, at best, rely on

a summary of papers deemed to be statistically significant versus not statistically significant. Again, using statistical significance is especially limiting as this approach is highly susceptible to sample sizes and again relies on crude categorizations of effective and ineffective, rather than discussing the magnitude of the relationship or conditions where a relationship is strengthened or weakened (Lipsey & Wilson, 2001).

The third rationale for conducting a meta-analysis is that the coding process itself brings the researcher and the reader into the inherent messiness in social science research where studies of different designs have varying results (Card, 2012). A key feature of meta-analysis is a focus on having the researcher identify the characteristics of the population studied, the features of the intervention, and the research design or analytic approaches used, which could help understand why results vary and which factors are associated with more positive or negative outcomes. Moreover, by including and ultimately combining studies of varying sample sizes, meta-analysis gains statistical power that enables the further examination of the variables that account for the variance in the results. The systematic approach can lead to the identification of both individual, environmental and study design factors that influence the magnitude and direction of the relationship examined (Bowman, 2012).

Finally, meta-analysis provided an organized approach at handling information from an incredibly large body of potential studies (Borenstein et al., 2009; Hunt, 1997). Systematic coding allows a level of detail and consistency that is challenging under other vote-counting or note-taking procedures. The meta-analysis approach also provides a rationale for what studies were considered and the criteria that lead some studies to be included and others to be precluded. Others may present a reasonable argument for why

the research should have included or excluded a study, but the detailed meta-analysis provides the roadmap for the reader to make meaning of the decisions made by the researcher.

Weakness and critiques of meta-analysis. While the above are reasons in support of using meta-analysis as an analytical technique, there are numerous critiques and purported weaknesses to the technique. Many of these critiques are valid; however, I suggest that many are in fact arguments that should be pointed towards all literature reviews in general and are general critiques of quantitative methods more broadly.

One critique is that that conducting a meta-analysis requires extensive statistical expertise (Card, 2012). At its most basic level, a meta-analysis is a series of ANOVA and regression analyses with additional consideration for study weighting, a level of expertise one could reasonably expect from a graduate level statistics course. Like any research approach, there are varying levels of complexity that can be explored and utilized. Given that meta-analysis is seldom a part of quantitative methods courses, researchers have argued that it is not the statistical expertise that is most lacking but the time and patience to code studies, and the ability to effectively communicate the research results (Card, 2012; Lipsey & Wilson, 2001).

A second critique is that highly structured approaches to meta-analysis lack the 'finesse" of qualitative or ethnographic approaches (Card, 2012). Like other critiques, this is one often levelled at quantitative methodologies more broadly. In response, many of the categories that will be coded in the current study in terms of population, study abroad program and research design characteristics emerge from existing qualitative research and more theoretical work. In terms of the coding process for individual studies,

it is like an approach used for coding interview and focus group transcripts (Lipsey & Wilson, 2001). Researchers could use a grounded theory approach in meta-analysis and code all aspects of possibly 10-20% of studies. Then the codes and the structure could be evaluated and reviewed by an external auditor for overlaps and consistency before proceeding with the rest of the studies. An alternate approach would have an *a priori* set number of categories and then code each study only on the characteristics identified at the onset of the research study. Both approaches could include the search for the presence or absence of certain characteristics. In fact, a qualitative review and a meta-analysis can be performed on the same exact studies.

A third and more common, consistent, and justified critique is the "apples and oranges problem". This dates to Eysenck's 1978 critique of the Smith and Glass's (1977) original meta-analysis publication wherein studies that used diverse approaches to psychotherapy treating a wide variety of problems in varying populations were used. Clearly, this critique has merit, but only to the extent that the researcher wants to draw conclusions about apples and oranges. A narrowly defined population of studies would only attempt to make conclusions based on those studies and populations. Other researchers, like Glass and Smith, looking at the broader topic of psychotherapy would recognize the variety of approaches, populations and challenges that are present in its practice. Their research would attempt to make conclusions about psychotherapy more broadly but could do so only if they included a very broad range of studies (or types of fruit). The apples and oranges critique may be better thought of in terms of sampling. The broader the sample of studies included in the analysis, the more widely the findings may be generalizable too. In the current study, I have not said that a study abroad program

must be a specific length, or taught in a foreign language, or led by home faculty. This is done to reflect the variety of ways that a student can study abroad and the study abroad research is conducted. Similarly, I use a very broad definition of intercultural competence, one that allows for the inclusion of many different instruments that measure the construct that study abroad program directors, institutional leaders and policy makers

Rather than viewing a broadly defined construct or relationship as a weakness, the inclusion of various types of studies presents an opportunity and underscores a distinct advantage of the meta-analysis over a traditional review of literature (Glass in Robinson, 2004). By analyzing studies with a variety of approaches, designs and population, the researcher can investigate how results are associated with those characteristics, also called moderators. The role of the meta-analysis researcher is not to identify the best designed research study, but to examine the range of studies that attempt to measure a given relationship. The decisions for meta-analytic researchers lie in deciding what characteristics to code in the set of available studies and the values they assign to each of them.

The fourth critique is publication bias. This refers to the possibility that the studies included in the meta-analysis are not reflective of all those that have been conducted (Card, 2012; Rosenthal, 1979; Rothstein, Sutton & Borenstein, 2006). The rationale underlying this concern is that journal editors are believed to be more likely to publish studies with statistically significant findings than studies that found no significant relationship between the variables of interest. However, there are several tests that can be run to identify the potential and magnitude of publication bias in a meta-analysis.

Study quality is the fifth critique and is also partially related to the apples to oranges argument from earlier (Lipsey & Wilson, 2001). This critique focuses on the issue that meta-analysis includes primary studies of varying rigour. The argument follows that a meta-analysis should include only studies of the highest quality and ignore those of lower quality. However, as argued by Glass (1976), the issue of study quality is both a coding task for the meta-analysis researcher, and a testable research question. Studies should be included if they meet the inclusion criteria developed before the study commences, not later.

Defining quality is clearly a problematic endeavour in academic research. A more appropriate approach may be including studies that meet the desired criteria but acknowledging the variance in quality, and in fact using the areas that studies vary on as a basis for additional research. In these cases, the researcher examines the relationship between *x* and *y* but looks to see how much of that relationship is impacted by, for example, the survey instrument used in the research project. This allows a test to see if survey instruments that are of lower quality do in fact lead to weaker, or opposite direction relationship between the variables of interest. Conducting this analysis allows the researcher to state which aspects of study quality are related to the observed effect sizes, which can serve as a guide for future researchers. Clearly, there are numerous challenges and benefits that a researcher is confronted with in choosing a meta-analytic approach. However, as with any valid and trustworthy piece of research, it is incumbent on the author to detail the decisions they are making in the research process to the reader so that they may understand the choices and critique the authors based on those decisions.

Meta-analysis in higher education. Over the past 30 years, meta-analysis has become a commonly used analytical approach in research as a growing number of researchers have seen its relevance in synthesizing large bodies of empirical work (Card, 2012). However, for researchers in education, especially higher education, meta-analysis has not been nearly as popular. Though its use in higher education is far from widespread, there is evidence of its use in this area of research. For example, Crede and Kieszczynka (2010) examined the relationship between attendance and academic achievement. Denson (2009) has conducted meta-analysis on the relationship between college diversity activities and the likelihood of racial bias in students. Bowman has further examined diversity experiences in looking at its relationship with cognitive development (2010) and civic engagement (2011). Several studies about college students and various outcomes related to health and wellness have been examined including physical activity behaviours (Keating, Guan, Piñero & Bridges, 2005), gambling (Nowak & Aloe, 2014) and changes in weight (Vella-Zarb & Elgar, 2009). Though meta-analyses remains rare in higher education literature, two of the most frequently cited texts in higher education, *How College Affects Students Volumes I, II* (Pascarella & Terenzini, 1991; 2005) and *III* (Mayhew, Rockenbach, Bowman, Seifert, Wolniak, Terenzini & Pascarella, 2016) also include meta-analytic results of some of the outcomes examined in the texts. Still, meta-analysis remains a very small proportion of all literature reviews and an even smaller proportion of research studies in higher education.

Searching the Literature for Relevant Studies

In his 2012 text, Card outlined a 12-step process for locating the studies that provide the answers to the desired research questions. The next section provides a streamlined review of that approach as it applies to the current study.

Developing and articulating the sample frame. Whereas the purpose of a literature review is to describe and critique a large amount of research that focuses on the construct(s) of interest, the purpose of the meta-analysis is to obtain and analyze the population of studies that exist on those constructs. In doing so, it is necessary to move beyond the traditional ideas of studies as only journal articles and other format including dissertations, master's theses, technical reports and conference presentations. This approach is broad and inclusive and is somewhat akin to thinking about sampling frames in survey research and approaches to minimize coverage error and minimize publication bias (Card, 2012). Coverage error in survey research occurs when "not all members of the population have a known, non-zero chance of being included in the sample and from those excluded differing from those included" (Dillman, 2009, p. 81).

The first type of error, undercoverage, is the largest concern in this study. It occurs when "there are elements in the target population that do not or can not appear in the sampling frame" (Dillman et al., 2009, p. 72). This is essentially the researcher's inability to get all potential studies that could be considered into their frame. In the proposed meta-analysis, undercoverage would be a situation where a study that meets the inclusion criteria is not included. To maximize the number of studies in the meta-analysis and obtain as close to a population of potential studies, a multi-step search procedure, outlined later, is used. Note that a study that does not include enough data to calculate

effect sizes could still be included in the meta-analysis, if these additional details can be obtained from the authors. A second coverage concern is ineligible units, also known as overcoverage. These are cases where a study is included that should not be (Dillman et al., 2009). To minimize this, strict criteria are used to determine study inclusion which are detailed in the next section.

A third example of coverage error is duplication. Duplication occurs when the same study is included twice in the meta-analysis. In the proposed meta-analysis, duplication could occur if a research study was presented first as a dissertation and later as a journal article. In these cases the publication which provides the most complete information is the one used for analysis. Additionally, if the same data set is used to publish multiple articles or by different authors, duplication could occur. To avoid this, I focussed on the researchers' names, samples sizes, survey instruments used and descriptions of data collection procedures (Dillman, et al., 2009). These are the steps that guided the study collection process, ensuring that as many studies as possible but no duplicates, were on the list of analyzed studies.

Criteria for inclusion. Narrative summary approaches to summarize the literature related to study abroad and intercultural competence can be found in effectively every article, dissertation or report on those two constructs. In this study I took a very broad approach to study inclusion. Each inclusion criterion is also a potential limitation to generalizability as it narrowed the scope of the study (Card, 2012). Still parameters are required, and four criteria form the basis for this meta-analysis.

First, the study must be written in English. Though study abroad is a global activity, any studies which are written in a language other than English are not included.

This approach ensured that research, regardless of the specific population and geographic area of a study abroad program studied, from around the world to be included, provided that the report was in English. The second inclusion criterion was that the paper must focus on a for-credit study abroad program. There are numerous other highly valuable international programs and activities, but they may not necessarily award academic credit. As noted earlier, the operational definition of study abroad is "defined as educational programs of study, in countries outside the students' home institution, that lead to academic credit" (FEA, 2011, p. 33).

Thirdly, the study must focus on intercultural competence. Though other outcomes associated with study abroad such as foreign language acquisition, personal or identity development, academic or subject knowledge, grade point average, graduation or retention rates are interesting, they are not the focus of this study. If a study focusses on intercultural competence and another outcome, then the study could be included though only the intercultural competence portion of the study would be a part of the meta-analysis. Finally, and strongly related to the third point, the study must present the intercultural measure quantitatively. If a study was mixed methods data collection, the quantitative section of the report could still be included. Within each individual study, I identified and coded as many student, study abroad program and research design characteristics from the article and follow up contact with the authors as possible.

Study collection procedures. In this meta-analysis, I used five primary strategies to identify potential studies (Card 2012; Lipsey & Wilson, 2001). The first strategy was a search of several academic databases including Educational Resources Information

Clearinghouse (ERIC), PsychINFO and Dissertation Abstracts. Keyword searches used the following set of search terms:

("stud* abroad" or "education abroad" or "international study" or "foreign student" or "student exchange" or "exchange program*" or "international educational exchange" or "Student Exchange Programs" or" Exchange Program*" or "Foreign Students;")

AND

("inter*cultural" or "cross*" or "global" or "cultural awareness")

AND

("college student*" or "undergraduate student*" or "university student*")

These keywords were developed from reviewing the keywords associated with some of the most commonly cited articles about study abroad and intercultural competence. Specific focus was paid to search terms that denoted the population of interest (college and university students) and context (study abroad). In addition, terms that approximated the names of the major survey instruments were included in the criteria

All APA citations and details from the databases were downloaded or copied into a reference manager and sorted for duplicates. The abstracts were reviewed to identify if they met the four criteria of being in English, focussing on intercultural competence with a quantitatively measured outcome of a for-credit study abroad program. Then, the studies were read, and discarded if it was later found that they did not meet the inclusion criteria.

The second step involved backwards and forward searches. Backwards searches were done by reviewing the reference sections of papers to identify older studies that did not emerge in the database searches (Card, 2012; Lipsey & Wilson, 2001). Forward

searches involve an approach based on papers that have cited already identified papers. This is enabled by using tools like Google Scholar. Whereas the backwards search is useful for identifying older published works, the forward approach helps identify newer, possibly less or uncited works. The forward approach was used on 15 of the most highly cited articles under the assumption that more recent relevant works would cite some of the more seminal articles from the field (Card, 2012).

A third search approach was a manual search of the table of contents of journals identified in the previous steps that have published relevant study abroad articles (Card, 2012; Lipsey & Wilson, 2001). The search included the following journals: *Journal of College Student Development, Journal of Student Affairs Research and Practice, Journal of International Education, Journal of Studies in International Education, Journal of Research in International Education, International Journal of Higher Education, the Frontiers the International Journal of Study Abroad, IDP Database of Research on International Education*, and other study abroad research bibliographies. These were reviewed by examining the table of contents, going back, where available to 1990 and abstracts for studies that potentially meet the inclusion criteria.

The next approaches were less structured. The first involved approaching individuals associated with some of the primary survey instruments used in study abroad research (as they may be aware of who has used their instruments in the past) and asking for bibliographies of those who have used their instrument. The final approach was a crowdsourcing effort. For this I shared my list of included studies through study abroad listservs, and study abroad and international education association emails and social

media using a Google drive document (linked here) asking for individuals to identify any potential studies they see missing from the list.

Each paper retrieved was recorded in MS Excel. Studies using the same dataset were identified, but the study providing the most complete details regarding the coding characteristics was maintained. An example of this is the Georgetown Consortium Study. The original published study (Vande Berg et al., 2009) was not used in this meta-analysis, as Nichols' (2011) dissertation provided more details than the original study, though the Vande Berg et al., study is cited throughout this dissertation. This database included the following categories; numerical ID assigned to study, author(s), year published or delivered, title of paper or presentation, abstract, which search procedure identified the study, and the step it was excluded from the study. This became the list of studies that were coded for the actual meta-analysis.

Data coding procedures. Data from each study included in the meta-analysis were coded into one of five categories. The first (see Appendix A) is bibliographic information, which is primarily data about the type of study, where it was published, the author and their institution. This section also included information about when the study was retrieved, when it was coded and if there was a second coder. The second category (see Appendix B) focuses on the student characteristics (Astin's 'I') in the study, both those who studied abroad and, if part of the study, those who did not. This section codes for the proportion of the participants by gender, program of study, year of study, international/domestic, ethnocultural/racial identity, and previous experience abroad. In these codes, I used the original data from the reports as much as possible. For program of study, business, commerce, finance, marketing were all considered "business". STEM

programs focussed on anything relating to physical, biological and life sciences. Arts and Social science comprise most of the rest of the programs of study. These were all coded as the percentage of students in a study who are identified in each category.

The third category (see Appendix C) looks at the characteristics of the study abroad program of the participants or (Astin's 'E'). This includes the program duration, language of instruction, context of program (who taught the courses), type of housing arrangements, guided opportunities for cultural interaction, opportunities for reflection. The fourth category (Appendix D) focusses on research design characteristics of the study itself. This includes the nature of the design (pre-post, pre-post with comparison group), instrument name, nested participants or not (single group, or individual students) and finally the single name construct of what outcome they were researching. With these variables coded, I began to enter values from the study directly into the Comprehensive Meta-Analysis software, the platform used in the study to conduct the meta-analysis. I coded the sample sizes, the pre-test and post-test mean scores for the abroad and at home groups, their standard deviations, as well as noting any additional statistics presented (*F*, *r* and *p*-values).

Trustworthiness of Coding. I was the primary coder for all analysis in this dissertation and reviewed and coded all studies meeting initial inclusion criteria. Identifying coded variables helps to increase transparency and replicability, two of the most important concepts in meta-analysis (Wilson, 2009). However, reliability of coding decisions is also an important aspect of the process.

To improve the reliability of the coding, two graduate students were recruited to be independent raters. Reviewer 1 reviewed 15 studies, reviewer 2 examined 11 studies.

They were compensated for their individual time reviewing studies and for the training time as well ($20 per hour). The training followed this format:

1. An introductory meeting where we reviewed the procedures of the study and the individual items to be coded in each section (2 hours).
2. One session where the raters and I completed the coding protocol together for 3 studies (1.5 hours x 3 studies = 4.5 hours).

After completing the training, the raters were given a few studies at a time, selected at random, from the studies that had been collected and coded at that point. The raters coded their studies and we I met with each of them separately to compare to our ratings on the selected studies. Table five contains the percent agreement for each major category.

Table 5

Agreement Rates of Coders

	Reviewer 1 (k = 15)	Reviewer 2 (k = 11)
Gender	100%	100%
Program of Study	87%	73%
Year of Study	100%	100%
Ethno-cultural Identity	100%	100%
International/Domestic	100%	91%
Previous Experience Abroad	100%	100%
Duration	87%	73%
Language used	87%	73%
Context	93%	82%
Housing	87%	82%
Cultural Interaction	33%	40%
Guided Reflection	53%	45%
Publication Type	100%	100%
Researcher Involvement	87%	82%
Survey Instrument	100%	100%
Research Design	100%	100%

Differences in program of study were typically around programs that might be considered a social science versus a STEM program. Differences in duration were due to questions

about duration of the study abroad program versus the length of time between the surveys. These differences were resolved during our meetings. The most significant discrepancy was with respect to cultural interaction and guided reflection. These differences were about which level to score on the Engle and Engle classification system. After much discussion with both students, I decided to change the coding structure for both from a scale of 1-5 to a simple yes or no. Yes indicating that there was some evidence of cultural interaction/experiential education or no there was not. Yes there was some evidence of guided reflection or no there was not. The intensity or magnitude of frequency was ignored. With the simplified variable there was full agreement with the students on the coding decisions.

General Procedures for Calculating Effect Sizes

The key point of conducting a meta-analysis is to understand not just if there is a relationship between two constructs or populations of a treatment, but the magnitude, the direction and the conditions that lead to larger, smaller relationships. The rest of this chapter outlines the analytical approach used to answer the research questions. All the equations presented here are from Borenstein et al. (2009). There are, however, numerous variations on the notation used in these equations (see Card, 2012, Lipsey & Wilson, 2001; Hunter & Schmidt, 2017 for others). Borenstein et al.'s notation was chosen as their formulas are used in the software, Comprehensive Meta-Analysis, which is designed specifically for meta-analysis.

To answer the research questions in this study requires a shift in focus from statistical hypothesis testing to effect sizes (Borenstein et al., 2009). Effect sizes are an "index of the direction and magnitude of association between two variables" (Card, 2012,

p. 87). The calculation of effect sizes allows for the comparison and synthesis of studies using a single metric. The effect size can be measured using many different statistics including correlations, differences between groups or contingencies between dichotomous variables (Borenstein, et al., 2009). While there are numerous effect sizes that could be used in this meta-analysis, Hedge's *g* was chosen for three reasons. First, though used only sparingly in study abroad research, it was the only effect size found in the studies used in this dissertation. This makes comparisons quite easy. Secondly, *g*, has a well known, though imperfect interpretation scale. Trivial effect sizes (+/-) are $0.0 < g < 0.20$, small effect sizes (+/-) $0.20 < g < 0.50$, medium effect sizes (+/-) $0.50 < g < 0.80$ and large effect sizes are anything larger than (+/-) 0.80. Card (2012) recommends using an effect size of standardized mean difference (like *g*) when one of the variables in the study is dichotomous. In this dissertation study abroad or studying at home is that dichotomous variable. Finally, some (see Card, 2012) have argued to use Hedge's *g*, rather than Cohen's *d* as it corrects for small sample sizes. Given that in this meta-analysis,10% of the studies have samples under 10, Hedge's *g* is justified as the choice for this study.

In meta-analysis, equations for Hedge's *g* are most often expressed initially as Cohen's *d*. The first is the definitional formula for Cohen's *d*, (Borenstein et al., 2009):

Equation 1 $$d = \frac{\bar{Y}_2 - \overline{Y_1}}{sd_{within}}$$

where Y_2 *and* Y_1 are means at times 2 (post-test) and time 1 (pre-test) and sd_{within} . In the cases that only the sd_{diff} is reported, equation 2 can be used to calculated sd_{within}

When working with dependent measures a formula for S_{within} is required

Equation 2 $$sd_{within} = \frac{sd_{diff}}{\sqrt{2(1-r)}}$$

A unique piece of equation 2 for dependent or repeated measures data is the correlation between the pre-test and post-test.

Equation 3 $$V_d = \left(\frac{1}{n} + \frac{d^2}{2n}\right) 2\,(1-r)$$

Where *n* is the number of matched pairs in a study.

Finally, the standard error of Cohen's *d* was computed using the formula below (Borenstein et al., 2009)

Equation 4 $$SE_d = \sqrt{V_d}$$

While the above shows the standard approach, there are numerous ways to calculate a study effect size and to convert between effect sizes. This is another benefit of using the Comprehensive Meta Analysis software as it can create an effect size based on numerous inputs. Using the software, rather than manually calculating formulas, drastically reduces the risk of human error. In this study, nine different approaches to calculating the effect sizes were used based on the information available in the study.

Studies with a comparison group. While not used all the time, some research on study abroad uses a comparision group. In order to calculate a mean effect size for the entire study, the effect sizes for both groups must be calculated using the formula below, starting by calculating *d* (Borenstein et al., 2009).

Equation 5 $$d = \left(\frac{(\bar{Y}_{2,T} - \bar{Y}_{1,T}) - (\bar{Y}_{2,C} - \bar{Y}_{1,C})}{SD_{pre+post}}\right)$$

The formula takes the above subtracts the difference of the pre-test from the post-test of the group at home (C) from the group abroad (T) and divides it by the pooled standard deviation from both measures. The pooled standard deviation is defined below.

Equation 6 $sd_{pre + post} = \left(\sqrt{\frac{(n_T-1)SD^2_{pre,T} + (n_C-1)SD^2_{pre,C} + (n_T-1)SD^2_{post,T} + (n_C-1)SD^2_{post,C}}{(2*(n_T\, n_C-2))}} \right)$

Once a Cohen's d has been calculated for a study, a correction is applied to convert to Hedge's g. In this conversion, *J* is commonly used, the degrees of freedom (df) come from the within estimate (*sd*).

Equation 7 $$J = 1 - \frac{3}{4df-1}$$

A positive effect size occurs when the diffference between the abroad group (T) is greater than the pre-post score difference for the at home group (C). Commonly, studies provided only descriptive statistics along with *p*-values and *t*-statistics when paired samples *t*-tests were used. In these situations, the effect sizes were calculated using the data presented. When additional data or statistics were required to compute the effect size, I contacted the primary author(s) of the study for these details. If the data could not be obtained, the study was excluded from the meta-analysis.

Correlation between pre and post-tests. For dependent groups studies, which are cases where the focus is only on pre- and post-test scores of study abroad participants, r, the correlation between pre- and post-test scores is needed (Card, 2012). the formula for r found below.

Equation 8 $$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

There are only a few cases where these figures are provided or can be computed from data in the study. In a few cases an actual pre-post correlation statistic was provided. In some small studies, the authors provide line item pre- and post-test scores for each participant, allowing for the calculation of *r*. Finally, in some cases where regression was used and the first step regressed the post-test against the pre-test, we can use the square

root of the adjusted R^2 squared to calculate the correlation. Where no correlation statistic was available one of two approaches was tried. For instruments with published psychometric data available, the test-retest statistic was used. For all other cases I used a correlation equal to 0.5 (personal correspondance Dr Steven Tarlow, 2018).

Fixed and random-effects models. In determining an analytical approach for the meta-analysis, the decision is primarily about choosing a fixed effect or random effects model for analysis. This decision should be made before the analysis begins, rather than beginning with a fixed effects model and if significant variance is found, changing to a random-effects model. Instead, researchers should make this determination based on the nature of the data and be guided by theory about the intervention and relevant variables. According to Card (2012), there are four considerations for making the decision about which approach to use. The first is the presence or absence of heterogeneity in the effect sizes; that is if it is reasonable to expect significant variance in the effect sizes beyond what would be expected within a normal sampling distribution a random effects model should be used. The second consideration is the amount of statistical power present in the meta-analysis. If there are less than 20 studies, a fixed-effect approach is likely better as random-effects models require additional statistical power. Given that over 80 effect sizes are used in the study, this is not likely a concern. A third consideration includes the presence or prevalence of outliers in the effect sizes in the study. If it is believed or known that there are outliers, a random effects model is advised. The presence of outliers can be identified using a visual graphing technique of the effect sizes like a stem and leaf plot which will be described later. The final consideration deals with external validity, that is, the degree to which you want to be able to generalize findings in the meta-

analysis. If a researcher is satisfied making conclusions based only on the studies used in the meta-analysis, then a fixed-effect model is preferred. However, to make broader generalizations, ones that extend to studies, populations, and study abroad programs not included in the meta-analysis (but still fitting the inclusion criteria), a random effects approach is warranted.

Random effects model. Given the number of characteristics under examination in this study, it is reasonable to expect significant variation in the mean study effect sizes justifying my decision to use a random effects approach. In addition, this study has very broad inclusion criteria and it is expected that there will be several dozen effect sizes, giving the study adequate statistical power. This study covers a broad range of studies that includes differing populations, program designs and research characteristics, which supports an approach that allows for broad generalizability.

Analysis of random-effects. In a fixed effect model, the assumption is that true effect size for all studies is identical and that the only differences are due to sampling fluctuations. However, in a random effects model, we are attempting to estimate a mean of the distribution of effects. A random-effects model has three components, while the fixed effects model has only two. These are shown in equations 9 and 10 in order.

Equation 9 $$Y_{i\ =} = \theta + \varepsilon_{i}$$

Equation 10 $$ES_i = \mu + \xi_i + \varepsilon_{i}$$

In equation 9 for a fixed effects model, all effect sizes in a study are assumed to be equal except for the sampling error $\varepsilon_{i.}$ By comparison in equation 10 the observed effect size of study is the sum of the mean of the population effect sizes (μ), plus the reliable (not due to sampling distribution) deviation of study *i* from the mean of the distribution of

population of effect sizes (ξ), plus the sampling deviation of study *i* from the distribution of population effect sizes (ε_i) (Borenstein et al., 2009).

The next step in using a random effects model is to identify weights for each study using the following formula

Equation 11 $$w_i^* = \frac{1}{\tau^2 + SE_i^2}$$

In Equation 11, SE_i^2 is the standard error of the effect size of study *i*, and τ^2 (tau-squared) the variance between studies. To calculate τ^2 requires an additional calculation found in equation 12

Equation 12 $$\tau^2 = \frac{Q-(k-1)}{(\sum w_i) - \frac{(\sum w_i^2)}{(\sum w_i)}}$$

The formula for tau-squared introduces *Q*, the heterogeneity statistic (see equation 13 for formula), *k*, the number of studies in the analysis, and w_i, the weight for an individual study. The new random effects weight for study *i* is a function of both the population variance and the standard error of study *i*. The addition of τ^2 means that the weights vary less in the random effects approach. This will result in a smaller sum of weights across the studies, and a correspondingly larger standard error. With the individual study weights, the next step is to use that estimate to calculate the *Q* statistic, which is the measure of heterogeneity

Equation 13 $$Q = \sum(w_i(ES_i - \overline{ES})^2)\ \sum\left(w_i ES_i^2\right) - \frac{(\sum(w_i ES_i))^2}{\sum w_i}$$

Equation 13 focuses in w_i, the weight assigned to study *i*, ES_i the effect size of study *i*, $\overline{ES}$ the mean effect size across all studies in the analysis. The resulting *Q*-statistic can then be compared to the χ^2 distribution with the corresponding degrees of freedom. If this

analysis allows us to reject the null hypothesis that the studies are homogeneous, it lends greater support to using the random effects model (Card, 2012).

The final three formulas return to the mean study effect size in presenting the standard error of the mean of the effect size, the significance test for the mean effect size (Z) and 95% confidence intervals.

Equation 14 $$SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w^*}}$$

Equation 15 $$Z = \frac{\overline{\mathrm{ES}}}{\mathrm{SE}_{\overline{\mathrm{ES}}}}$$

Equation 16 $$ES_{CI} = \overline{ES} \pm Z_{1-\alpha} SE_{\overline{ES}}$$

In equation 16, to calculate 95% confidence intervals a *Z* value of 1.96 would be used. These are the formulas used by the Comprehensive Meta-Analysis Software to conduct the analysis in this study.

Research Question 1: Does an overall relationship exist between study abroad participation and intercultural competence? Is there significant variation in this relationship across studies?

The previous section outlined the effect size that will be used throughout the meta-analysis, the approach for calculating the standard error of each study effect size, and the procedure that will be used to weight each study. These steps are precursors to the computation of the mean effect size of the relationship between study abroad participation and intercultural competence. Calculating the weighted mean effect size was done using the following equation (Borenstein et al., 2009)

Equation 17 $$\overline{ES} = \frac{\sum(w_i^* ES_i)}{\sum w_i^*}$$

$\overline{ES}$ = mean effect size
w_i^* is the weight for study i

ES_i is the effect size for study i
$\sum w_{i*}$ is the sum of the weightings

The analysis from the previous equations provide the answer to research question 1: What is the mean effect size and the 95% confidence intervals among the studies examining the relationship between study abroad and intercultural competence? This analysis provides only a response to the question of average (weighted) effect size. It does not identify which factors are associated with variations in the mean effect size.

The analysis to answer the first part of research question one, "Does an overall relationship exist between study abroad participation and intercultural competence?" comes from calculating the mean effect size, its standard error, the 95% confidence intervals and the significance test of the mean effect sizes. The analysis for the second part of the first research question, "Is there significant variation in this relationship across studies?" is answered by completing the heterogeneity and population variability estimates. The next section focuses on the second research question which attempts to identify if, and to what extent, individual study characteristics (student, study abroad program, and research design) can explain variation in mean study effect size.

Research Question 2: To what extent do individual student, study abroad program and research design characteristics moderate the relationship between study abroad and intercultural competence

The second research question is effectively three research questions in one. Part one focuses on extent to which individual student characteristics moderate the mean effect size. Part two focuses on the extent to which individual study abroad program, or environmental characteristics moderate the mean effect size. The final part examines the extent to which individual research design characteristics moderate the mean study effect sizes. The process for answering each of those three questions is effectively the same and

in meta-analysis is typically called moderator analysis (Card, 2012; Lipsey & Wilson, 2001).

As noted in the appendices there are many individual attributes that were coded in each section. The first step was to examine the relationships in each of these three sets of variables with mean effect size to determine which of them are most likely useful to include in moderator and meta-regressions.

Recall that Astin's model suggested that outputs are a function of inputs and the environment. Translating this to the planned meta-analysis, I was interested in how characteristics of the students who study abroad (Appendix B), and the characteristics of their study abroad program (Appendix C) are associated with the development of intercultural competence. In this study, the coded characteristics of the study abroad programs are largely from Engle and Engle's (2003) study abroad classification system. To this model, I also included an additional category specifically examining research design characteristics (primarily in Appendix D, though partially in Appendix A).

Research Question 3: To what degree do combinations of student, study abroad program and research design characteristics, and all characteristics explain variance in the mean study effect size?

The final analysis used variables from each of the three previous analyses and combined them into four separate meta-regression models.

Summary. The rationale for creating three separate but related questions is based on the idea that a given stakeholder typically only has influence over the decisions at one of these levels. A study abroad program advisor may be the one that chooses the students to go abroad, the one that sets the criteria, the one who decides if previous experience, or program of study are important characteristics in determining who studies abroad. The

study abroad program manager or faculty who leads the program the one who oversees, or is potentially considering developing a study abroad program, has the most influence over how their program is designed. They determine the duration, the nature of the academic coursework, and the possible housing options for participants. Finally, the researcher is not likely able to influence the students who decide or are interested in studying abroad, nor the specific features of the study abroad program. The researcher does however have influence on the way the study is conducted, whether a comparison group is used, and the instrument used to measure intercultural competence. These practical implications are strongly aligned with Astin's IEO model. All three parts of the model are considered when attempting to identify if there is a relationship between two elements of the college experience.

Conclusion

This chapter consisted of three sections. The first was an introduction to meta-analysis. This focused on the history and development of meta-analysis as a research technique. In this section I also focused on some of the rationales that support its use in research, along with some of the primary critiques or weaknesses too. The second part examined the process of searching and identifying studies that could be included in the study. This includes the inclusion criteria, the sources and approaches used to first identify studies. Following that I described the framework for how the studies were coded and the process I used to ensure reliable coding. The final part details the process for answering the three research questions.

Chapter 5: Descriptive & Inferential Results

Chapter five is structured around the findings related to the three research questions. This chapter outlines the results of the search process, the number of studies included, the number excluded on each criterion, tests of publication bias and the analysis relating to the three primary research questions.

Search Results

The search results produced 5745 possible studies published on or before December 21, 2016. Table 6 presents the results and the reasons and number of studies excluded for each reason.

Table 6

Identification and Attrition of Studies

Total unique sources identified		5745
Database searches	4757	
Forward and backwards searches	525	
Journal hand searches	300	
International Education Database hand search	156	
Internet & Community	7	
Rejected because study not in English		16
Rejected - Study could not be found, gated		21
Rejected - Not focused on study abroad		4656
Rejected - Duplicate		248
Rejected - Did not use intercultural competence outcome		235
Rejected - Dd not use not quantitative measures		274
Rejected - Did not use pre-test/post-test design		190
Rejected - Too few student, program, research design details		33
Included studies		72

Of the 5745 studies, sixteen were not in English and could not be read. Twenty-one were only abstracts and full versions of the study could not be obtained. 4656 were rejected on criterion one after reading the title and abstract for not focussing primarily on study abroad of university or college students. Two-hundred and forty-eight were found to be

duplicates. This included actual duplicates, plus dissertations that became articles and data sets that were used to publish more than one article with the same dataset. Twenty one studies could not be found. The third criterion was that the study used one or more intercultural competence outcomes as one outcome in their study. Two hundred and thirty-five studies were rejected on this criterion. The fourth criterion was that the study must have reported the intercultural competence outcome quantitatively – this led to the rejection of two hundred and seventy-four studies. 190 failed on the fifth criterion: that the studies must have a research design where student participants were surveyed at least two times on the intercultural competence outcome. 33 studies were rejected when not enough details could be extracted from the study, leaving 72 studies that lead to 85 unique effect sizes for analysis. A few studies included the results of multiple study abroad programs and provided enough details for them to be analyzed as individual effect sizes. These 85 effect sizes are outlined in table 7. The primary author, year of publication, instrument name(s), sample size, publication type, duration of study abroad program and study effect size (*g*) are shared.

Table 7

Summary of Included Studies

Primary Author	Year	Instrument Name	Sample Size	Research Design	Duration (weeks)	Effect size (*g*)
Anderson	2006	IDI	16	p-p	4	0.78
Anderson	2011	GPI & IDI	39	p-p	14	0.59
Armfield	2004	ISS	60	Retro	14	0.83
Bates	1997	GMS	49	p-p-c	14	0.75
Black	2006	CCAI	26	p-p	4	0.68
Black	2014	GMS	46	p-p	3	1.01
Boatler	1992	WMS	45	p-p	3	0.20
Burrow	2010	GPI	98	p-p	15	0.11
Burton	2013	IDI	9	p-p	3	0.02
Caffrey	2005	CC	32	p-p-c	5	1.86

Christopher	2008	IDI	27	p-p-c	4	0.34
Day-Vines	1998	IDI	24	p-p-c	6	-0.16
DeLoach	2015	GA	278	p-p	4	0.16
Edwards	2009	CCAI	32	p-p	4	0.83
Engberg	2015	GPI	510	p-p	14	0.32
Fabregas	2009	IDI	18	p-p	16	0.46
Fairchild	2006	MCA	15	p-p	2	0.55
Fernández	2006	GPI	219	p-p	4	0.12
Gaia	2015	GPI	136	p-p	3	0.15
Gingerich	1998	MGUDS	67	p-p-c	14	1.77
Hansen	2010	Other & GMS	51	p-p-c	10	-0.15
Harvey-A	2013	IDI	5	p-p	14	0.34
Harvey-B	2013	IDI	11	p-p	14	0.45
Hoff	2005	IDI	20	p-p-c	14	0.02
Hughes	2003	CCAI	47	p-p-c	10	0.68
Hunley	2008	CL	21	p-p	9	0.08
Hunley	2008	CL	46	p-p	8	0.16
Hyndman III	2009	CI	307	p-p	12	0.24
Jackson	2008	IDI	13	p-p	5	0.52
Johnson	2008	IDI	10	p-p	6	-0.09
Kafka	1967	WMS	81	p-p	14	0.09
Keefe	2008	IDI	39	p-p	2.5	-0.13
Kitsantas	2004	CCAI	232	p-p	5	0.50
Kitsantas	2001	CCAI	24	p-p-c	3	1.89
Kurt	2013	GA	30	p-p	4	0.17
Lemmons	2013	IDI	12	p-p	5	0.36
Lemmons	2013	IDI	18	p-p	4	-0.08
Lemmons	2013	IDI	9	p-p	6	0.27
Lombardi	2011	IL	13	p-p	6	0.40
Luchesi	2014	GPI	369	p-p	3	0.07
Maharaja	2009	CCAI	51	p-p	14	0.44
Mapp	2012	CCAI	80	p-p	2	0.62
Medina-Lopez-A	2004	IDI	16	p-p	7	0.08
Medina-Lopez-B	2004	IDI	9	p-p	16	0.37
Nam	2011	IDI	21	p-p	3	0.27
Nam	2011	IDI	18	p-p	3	0.16
Ng et al.,	2012	IDI	16	p-p	4	0.30
Nguyen	2015	IDI	55	p-p	3	0.63
Nichols	2011	IDI	1297	p-p-c	18.5	0.16
Ogden	2009	GC	219	p-p-c	12	0.21
O'Reilly	2010	SCAS	33	p-p	6	0.34

Pachmayer	2014	CC	174	p-p	14	0.25
Paige	2004	IDI	86	p-p	14	0.38
Palmer	2009	SCAS	93	retro	20	1.56
Patterson	2006	IDI	63	p-p-c	3	0.36
Pedersen	2010	IDI	29	p-p-c	35	0.01
Rexeisen	2009	IDI	19	p-p	16	0.83
Rexeisen	2008	IDI	39	p-p	14	0.06
Rexeisen	2013	IDI	126	p-p	14	0.23
Reza	2015	IDI	21	p-p	14	1.53
Roller	2012	IDI	19	p-p-c	13	0.14
Rust	2013	IDI	60	p-p	10	0.81
Sakurauchi	2014	IDI	8	p-p-c	10	1.02
Savicki	2004	ICAPS	65	p-p-c	14	-0.35
Savicki	2013	SCAS	27	p-p	14	0.27
Shaheen	2004	IDI	37	p-p-c	9	0.36
Smart	2014	GPI	174	p-p-c	13	0.26
Smith	2013	GA	53	p-p	2	1.38
Smith	2012	ISD	51	p-p	3	0.48
Spenader-A	2015	IDI	12	p-p	14	0.84
Spenader-B	2015	IDI	21	p-p	14	0.15
Spenader-C	2015	IDI	18	p-p	14	0.36
Spenader-D	2015	IDI	14	p-p	14	0.76
Spenader-E	2015	IDI	13	p-p	14	0.10
Spenader-F	2015	IDI	10	p-p	14	0.05
Spenader-G	2015	IDI	22	p-p	14	0.65
Spenader-H	2015	IDI	11	p-p	14	0.57
Stallman	2009	IDI	35	p-p	17	0.36
Stromberger	2011	GC & ISS	22	p-p	3	0.29
Teranishi	2008	CCAI	11	p-p	4	0.27
Tucker	2014	GE	54	p-p	2	0.73
Williams	2002	ISI & CCAI	52	p-p-c	16	0.61
Wortman	2002	OD	188	p-p-c	15	0.11
Zarges	2016	IDI	376	p-p	3	0.60
Zhai	2000	Other	98	p-p-c	5	0.28

Notes: ADS= Attitudes towards Diversity Scale, CC= Cultural Competency Scale, CCAI= Cross-Cultural Adaptability Inventory, CI=Cultural Intelligence, CL=Cultural Learning, GA=Global Awareness, GC=Global Competence, GE=General Ethnocentrism, GMS=Global Mindedness Scale, GPI= Global Perspectives Inventory, ICAPS= Intercultural Adaptability Potential Scale, IDI=Intercultural Development Inventory, IL=Intercultural Learning, ISD=Intercultural Skill Development, ISS=Intercultural Sensitivity Index, MCA=Multi-Cultural Awareness, MGUDS= Miville-Guzman

Universality-Diversity, OD=Openness to Diversity, SCAS= Sociocultural Adaptability Scale, WMS=World-Mindedness Scale.

In these studies, there were 6394 unique study abroad participants plus 640 in comparison groups. The average study in this dissertation had 75 study abroad students in its sample and a median sample of study abroad students of 26. For the 18 studies with a comparison group the average size was 30 and a median of 24. The next section focuses on the coded characteristics that were analyzed in the study.

Descriptive Data – Student Characteristics

Table 8 shows the coded student characteristics. Note the unit of analysis is the study itself, not the students, so the means should be interpreted as the average proportion of students with a given characteristic. In the table, *k* represents how many unique effect sizes included a valid measure of that 'x' in all the coded effect sizes.

Table 8

Description of Student Characteristics

Variable	*k*	Mean Percentage	SD	Low Percentage	High Percentage
Percentage Female	85	68%	0.14	36%	100%
Percentage Business	85	17%	0.33	0%	100%
Percentage Social Science	85	38%	0.41	0%	100%
Percentage STEM	85	13%	0.27	0%	100%
Percentage Other program	85	32%	0.41	0%	100%
Percentage 1st year	85	3%	0.07	0%	39%
Percentage 2nd year	85	15%	0.24	0%	100%
Percentage 3rd year	85	41%	0.32	0%	100%
Percentage 4th year +	85	22%	0.26	0%	100%
Percentage Other years	85	16%	0.34	0%	100%
Percentage Graduate students	85	3%	0.13	0%	100%
Percentage Asian & Pacific Islander	85	5%	0.19	0%	30%
Percentage African American	85	2%	0.05	0%	100%
Percentage Hispanic &/or Latinx	85	5%	0.18	0%	100%
Percentage White	85	41%	0.41	0%	100%
Percentage Unknown identity	85	47%	0.46	0%	100%
Percentage Domestic students	56	94%	0.19	0%	100%

Percentage International students	56	6%	0.10	0%	100%
Percentage Previous abroad	57	47%	0.29	0%	100%
Percentage Never abroad	57	53%	0.30	0%	100%

In the mean study, 68% of the participants were female and 32% male. Note to the end of the inclusion period (December 21, 2016) no studies identified any participants who chose an identity other than male or female. Among year of study, the mean study had 3% first year, 15% second year, 41% third year, 22% fourth year, 16% other or undefined and 3% graduate students. In terms of identity, the average study has 41% students who were identified as White, 47% did not have an ethnocultural/racial identity shared. Five percent were identified as Asian and Pacific Islanders, 5% Hispanic and/or Latino, 2% Black or African-American. Two additional variables, percentage of domestic students and percentage of students with previous experience abroad were only coded for 56 and 57 studies respectively.

Table 9 uses data from the Open Doors survey (IIE, 2017) to attempt to present an understanding of how representative the participants in this study are to those of U.S study abroad. Recall, the studies in this dissertation cover a much wider time frame than the available Open Doors data below. Furthermore, Open Doors is a voluntary survey of U.S institutions and does not purport to capture all participants.

Table 9

Study Abroad and U.S. Post-Secondary Enrollment

Variable	*U.S. Students Abroad 2006-07*	*U.S. Student Abroad 2015-16*	*U.S 2015-16 Post-Secondary Enrollment*
Percentage Female	66%	67%	56%
Percentage Business	19%	21%	16%
Percentage Social Science	31%	31%	27%
Percentage STEM	17%	26%	20%
Percentage Other Programs	33%	22%	37%
Percentage 1st year	4%	4%	

Percentage 2nd year	13%	13%	
Percentage 3rd year	34%	33%	
Percentage 4th year	20%	27%	
Percentage Unknown year	15%	9%	
Percentage Graduate students	10%	12%	
Percentage Asian & Pacific Islander	6%	8%	7%
Percentage African American	4%	6%	14%
Percentage Indigenous	1%	1%	1%
Percentage Hispanic &/or Latinx	5%	10%	17%
Percentage White	83%	71%	58%
Percentage Unknown identity	1%	3%	0%

The percentage of students identifying as female is quite similar, as is the year of study, though the Open Doors data reports a higher percentage of graduate students. Business and social science participation is similar but Open Doors reports a higher percentage of students in STEM program. Ethno-cultural identity is harder to compare as a large percentage of students were not identified in the studies used in this dissertation. Note that in the studies coded in this study less than 1% of students identified as Indigenous so the variable will not be used in later analysis. Open Doors does not track data relating to previous experience abroad nor the percentage of students who were domestic or considered international students. Moreover, Open Doors data presented in table 9 originate from two specific points in time and the studies in this dissertation cover a wide range of years than data is available.

Descriptive Data – Study Abroad Program Characteristics

Table 10 shows the six categories of study abroad program characteristics coded in the study that are measured categorically. The other variable, duration of study abroad program is measured continuously, and the average program duration was 9.4 weeks.

Table 10

Description of Study Abroad Program Characteristics

Variable	*k*	% of studies
Language of Instruction		
Primary language	42	49%
Mostly primary language	29	34%
Primarily foreign language	14	16%
Faculty Instruction		
Program taught by home professors	33	39%
Program taught by combination	40	47%
Program taught by local professors	12	14%
Housing		
Combination or not known	33	39%
Homestay	18	21%
With study abroad students	26	31%
With local students	8	9%
Nature of Study Abroad Program		
Single group or combination of groups	56	66%
Individual students and small groups	29	34%
Cultural Interaction/Experiential Education		
Yes	57	67%
No	28	33%
Guided Reflection		
Yes	48	56%
No	37	44%

Unlike table 9, the coding for study abroad program characteristics was done for the entire study. The '*k*' refers to the number of studies that had that characteristic, followed by what percentage that represents. The table highlights that half of the studies had students in programs entirely in a student's primary language (overwhelmingly English), about 1/3 mostly in their primary language (except for introductory second language) and the remainder were primarily or entirely in another language. For faculty instruction, 39% were taught only by home professors, 14% by local professors and the rest were done by some combination of the two. Student housing details found that about 21% were

homestay, 31% with other study abroad and international students, 9% with local students and 39% were a combination or not known. Study abroad duration is the only continuous variable in this section with program ranging from 2 weeks to a full academic year (considered 35 weeks in this study). The average is 9.4 weeks. Open Doors only calculates the duration of the programs and in 2005-06 the average was about 10 weeks and in 2015-16 it was down to about 8 weeks, thus our estimate falls within that range.

More than 60% of the studies were of a single group or set of groups of students, while the other studies were primarily studies of individual students. A way to think about this is that the former is most likely research done of a single group of students possibly in a faculty-led program, while the latter is an institution looking at the experiences of many of their students enrolled in study abroad during a specific time frame, possibly many bi-lateral exchange agreements. The two final characteristics were intentional cultural interaction and guided reflection. Two-thirds of studies indicated that participants engaged in cultural interaction or experiential learning, while just over half included guided reflection.

Descriptive Data - Research Design Characteristics

Research design characteristics were the last set of coded characteristics. They are indicators of the qualities and quality of the studies themselves.

Table 11

Description of Research and Study Design Characteristics

Variable	*k*	% of studies
Publication Type		
Article	41	48%
Dissertation	44	52%
Researcher Involvement in Study Abroad Program		
Yes	38	45%
No	47	55%

Year Published		
Before 1990	1	1%
1990's	4	5%
2000-2005	14	16%
2006-2010	24	28%
2011-2016	42	49%
Research Instrument		
Global Perspectives Inventory (GPI)	6	7%
Intercultural Development Inventory (IDI)	42	47%
Cross-Cultural Adaptability Inventory (CCAI)	9	10%
Global Mindedness Scale (GMS)	3	3%
Others	29	33%
Research Design		
Pretest-Post-test (P-P)	62	72%
Pre-post with control group (P-P-C)	21	25%
Retrospective pre-test	2	2%

About half of the studies identified were articles (including two conference papers) and half dissertations (including two master's thesis). The second variable was whether the faculty member(s) who authored the study were also participants in the study abroad program. Nearly half of the studies had the researcher involved in the program. The next characteristic is the age of the study. This will not be a studied variable; however, the chart shows the breakdown of included studies with about half the studies included from the current decade and the oldest being from 1968. The next characteristic is the instrument used in the research. Unlike all other variables in the study, this category uses $k = 89$ as four studies (Anderson et al., 2011; Hansen, 2010; Stromberger, 2011 & Williams, 2002) used more than one instrument. Almost half of all the effect sizes used the Intercultural Development Inventory, 10% the CCAI and 7% the GPI with about 1/3 using other instruments. Finally, 72% of studies used pre-test/post-test (pp) research designs, 26% pre-test/post-test with comparison group (ppc) and two used retrospective pre-tests.

Research Question 1: Does an overall relationship exist between study abroad participation and intercultural competence? Is there is there significant variation in this relationship across studies?

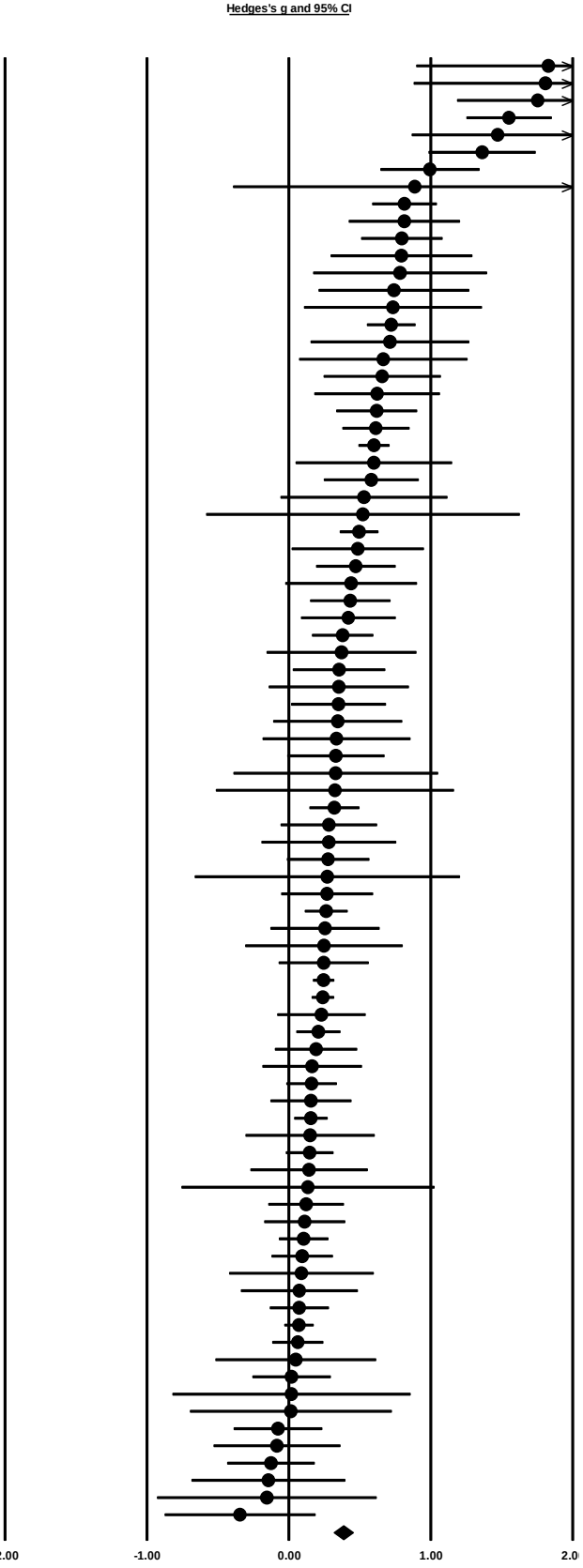
In the 72 studies, there were 85 usable effect sizes. They range from the smallest effect size, -0.33 (Savicki et al., 2004), which would be a small negative effect, to 1.83 (Kitsantas et al., 2001), which is a very large positive effect. Six of the effect sizes were negative, 79 were positive. In a fixed-effects model (Equation 9) the weighted mean effect size was positive and statistically significant $g = 0.31(.01)$, $Z(84)= 23.18$, 95% *CI* 0.29 to 0.34, $p < .001$. In a random-effects model (Equation 10) the effect size was slightly larger $g = 0.38(.03)$, $Z(84)= 11.13$, 95% *CI* 0.32 to 0.45, $p < .001$ and the standard error, as described previously, was larger. The random-effects approach indicates that the difference in intercultural competence between the pre-test and post-test in study abroad programs is about halfway between and small and medium size effect. Figure one provides a full illustration of each study effect size, the study effect size, the study standard error, plus the 95% confidence intervals for each estimate, ordered from the largest to the smallest.

Figure 1

Sorted Forest Plot of Studies and Effect Sizes

Summary of Studies and Effect Sizes

Study name	Outcome	Subgroup within study	Hedges's g	Standard error
Kitsantas et al. (2001)	CCAI	ALL	1.83	0.48
Caffrey et al. (2005)	CCHS	ALL	1.81	0.47
Gingerich (1999)	MGUDS	ALL	1.75	0.29
Palmer (2009)	SCAS	ALL	1.55	0.15
Reza (2015)	IDI	ALL	1.47	0.31
Smith et al. (2012)	Other	ALL	1.36	0.19
Black et al. (2014)	GMS	ALL	0.99	0.18
Sakurauch (2014)	IDI	ALL	0.89	0.65
Armfield (2004)	IDI	ALL	0.81	0.12
Edwards (2009)	CCAI	ALL	0.81	0.20
Rust et al. (2013)	IDI	ALL	0.80	0.15
Rexeisen et al.(2009)	IDI	ALL	0.79	0.25
Spenader et al. (2016)A	IDI	ALL	0.78	0.31
Anderson et al. (2006)	IDI	ALL	0.74	0.27
Bates (1997)	GMS	ALL	0.73	0.32
Tucker et al. (2014)	Combined	Combined	0.72	0.09
Spenader et al. (2016)D	IDI	ALL	0.71	0.29
Hughes (2003)	CCAI	ALL	0.67	0.30
Black et al. (2006)	CCAI	ALL	0.66	0.21
Spenader et al. (2016)G	IDI	ALL	0.62	0.23
Nguyen (2015)	IES	ALL	0.62	0.15
Mapp (2012)	CCAI	ALL	0.61	0.12
Zarges (2016)	MCCII	ALL	0.60	0.06
Williams (2002)	Combined	ALL	0.60	0.28
Anderson et al. (2011)	Combined	ALL	0.58	0.17
Spenader et al. (2016)H	IDI	ALL	0.53	0.30
Fairchild et al. (2006)	MC	ALL	0.52	0.56
Kitsantas (2004)	CCAI	ALL	0.49	0.07
Jackson (2008)	IDI	ALL	0.49	0.24
Smith et al. (2013)	Other	ALL	0.47	0.14
Fabregas Janeiro (2009)	IDI	ALL	0.44	0.24
Maharaja (2009)	CCAI	ALL	0.43	0.14
Harvey (2013)b	IDI	2.00	0.42	0.17
Paige et al. (2004)	IDI	ALL	0.38	0.11
Lombardi (2011)	Open	ALL	0.37	0.27
Shaheen (2004)	IDI	ALL	0.35	0.17
Patterson (2006)	IDI	ALL	0.35	0.25
Stallman (2009)	IDI	ALL	0.35	0.17
Spenader et al. (2016)C	IDI	ALL	0.34	0.23
Lemmons (2013)A	IDI	1.00	0.34	0.27
O'Reilly et al. (2010)	Other	ALL	0.33	0.17
Medina-Lopez-Portillo (2004)B	IDI	ALL	0.33	0.37
Christopher Brooks (2008)	IDI	ALL	0.33	0.43
Engberg et al. (2015)	GPI	ALL	0.32	0.09
Zhai (2000)	Other	Combined	0.28	0.17
Ng et al. (2012)	IDI	ALL	0.28	0.24
Stromberger (2011)	Combined	Combined	0.28	0.15
Harvey (2013)a	IDI	1.00	0.27	0.48
Savicki et al. (2013)B	SCAS	Combined	0.27	0.17
Smart (2014)	GPI	ALL	0.26	0.08
Nam (2011)A	IDI	ALL	0.26	0.20
Teranishi et al. (2008)	CCAI	ALL	0.25	0.28
Lemmons (2013)C	IDI	1.00	0.25	0.16
Pachmayer (2014)	Combined	Combined	0.24	0.04
Hyndman (2009)	Other	Combined	0.24	0.04
Rexisen et al. (2011)	IDI	ALL	0.23	0.16
Ogden (2010)	Combined	Combined	0.21	0.08
Boatler (1992)	WMS	ALL	0.19	0.15
Kurt (2013)	Other	ALL	0.16	0.18
Nichols (2011)	IDI	ALL	0.16	0.09
Hunley (2008)B	Other	2.00	0.16	0.15
DeLoach et al. (2015)C	gen CC	1.00	0.16	0.06
Nam (2011)B	IDI	ALL	0.15	0.23
Gaia (2015)	GPI	ALL	0.15	0.09
Spenader et al. (2016)B	IDI	ALL	0.14	0.21
Roller (2012)	IDI	ALL	0.13	0.45
Fernández (2006)	Other	ALL	0.12	0.14
Wortman	Open 2 Div	ALL	0.11	0.15
Burrow (2010)	GPI	ALL	0.10	0.09
Kafka (1968)	Other	ALL	0.09	0.11
Spenader et al. (2016)E	IDI	ALL	0.09	0.26
Hunley (2008)A	Other	1.00	0.07	0.21
Medina-Lopez-Portillo (2004)A	IDI	ALL	0.07	0.11
Luchesi (2014)	GPI	ALL	0.07	0.05
Rexisen (2013)	IDI	ALL	0.06	0.09
Spenader et al. (2016)F	IDI	ALL	0.05	0.29
Burton (2012)	IDI	ALL	0.02	0.14
Hoff (2005)	IDI	ALL	0.02	0.43
Pedersen (2010)	IDI	2.00	0.01	0.36
Lemmons (2013)B	IDI	1.00	-0.08	0.16
Johnson et al. 2008	IDI	ALL	-0.08	0.23
Keefe (2008)	IDI	ALL	-0.13	0.16
Hansen (2010)	Combined	ALL	-0.14	0.28
Day-Vines (1998)	IDI	ALL	-0.15	0.39
Savicki et al. (2004)	ICAPS	ALL	-0.34	0.27
			0.38	0.03



Heterogeneity in effect sizes. The next step is to examine the heterogeneity in the studies. The *p*-value for $Q = 404.75$ with $df = 84$ is less than .001. Using .05 as the alpha for statistical significance, the null hypothesis is rejected that all studies share a common effect size and accept the alternative that the mean study effect size varies among studies. The next step is to use equations 7, 8 and 9 to find estimates for $\tau^{2,}$ (the between studies variance) τ (between studies standard deviation), and I^2 (ratio of true heterogeneity to total observed variation) The estimate of τ^2 is .06 and the estimate of τ is .25. The tau-squared estimate is used to assign weights to each effect size in the random effects model. The significant result for *Q* suggests that it is highly unlikely that the 85 effect sizes do no share a common effect size. The formula for $I^2 = (Q - df)/Q = 79.25\%$, suggests that 79% of the variability is not due to sampling variability. A generally accepted rule is that an I^2 value greater than 75% represents considerable variance between-study effects (Higgins & Green, 2011). Thus, it is appropriate to move forward using a random effects model as these tests have highlighted significant heterogeneity. Individual moderators will be examined in research question 2 and groups of moderators in research question 3, but the next step is to investigate potential outliers and publication bias.

Visual displays of effect sizes. Before beginning analysis, it is helpful to visualize the unweighted effect sizes to have a sense of their distributions and potential outliers. The first approach is a stem and leaf plot which groups effect sizes. Figure two is organized in increments of g = 0.10, beginning with -0.3 and ending with 1.8. Each effect size is rounded to two decimal points and the hundredth point is plotted on the lines.

Figure 2

Stem and Leaf Plot of Study Effect Sizes

Stem													
-0.3	4												
-0.2													
-0.1	3	4	5										
-0.0	8	8											
0.0	1	2	2	5	6	7	7	7	9	9			
0.1	0	1	2	3	4	5	5	6	6	6	6	9	
0.2	1	3	4	4	5	5	6	6	7	7	8	8	8
0.3	2	3	3	3	4	4	5	5	5	7	8		
0.4	2	3	4	5	9	9							
0.5	2	3	8										
0.6	0	0	1	2	2	6	7						
0.7	1	2	3	4	8	9							
0.8	0	1	1	9									
0.9	9												
1													
1.1													
1.2													
1.3	6												
1.4	7												
1.5	5												
1.6	5												
1.7													
1.8	1	3											

Though there are six studies, with effect sizes $g > 1.30$ (Kitsantas et al., 2001; Caffrey et al., 2005; Gingerich, 1999; Palmer, 2009; Reza, 2015; Smith et al., 2012) that could be potential outliers. However, a review of them did not find coding errors, or similar qualities that might suggest they are from outside the population of studies that should be included in this analysis. Of those six studies, three were dissertations and three were

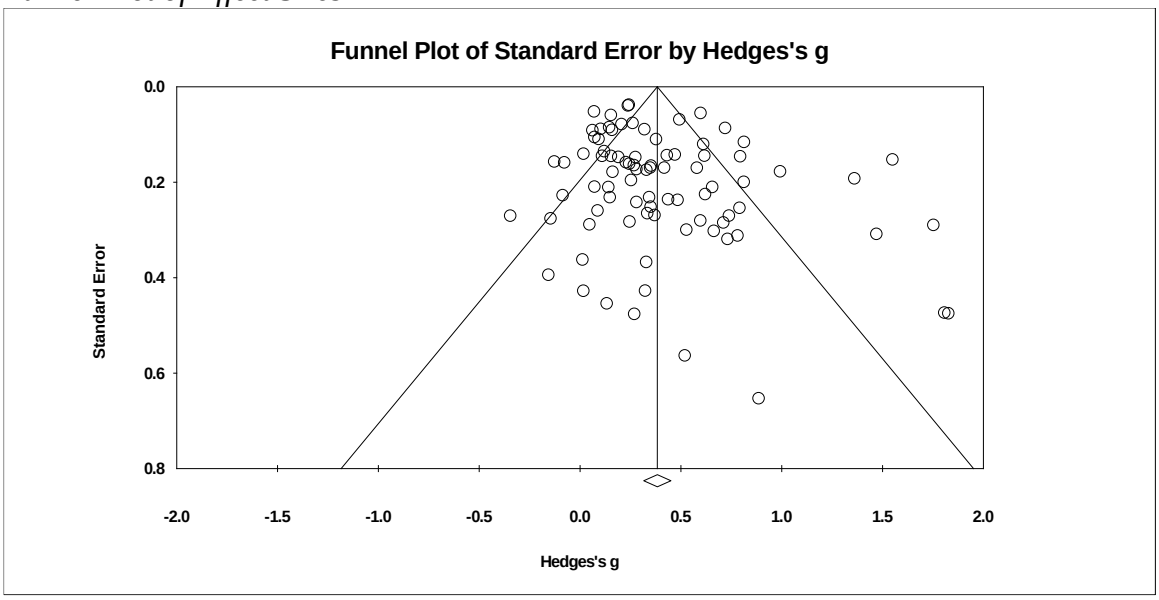
articles. One was a retrospective pre-test, while three used pre-post, and 2 pre-post comparison designs. The one quality that might make these studies unique is the research instrument used. While one used the IDI and another the CCAI, the other four studies used much less commonly used or personally developed instruments for their study. In addition, these studies, combined account for only about 6% of the total study weights which suggests they are not overly weighted in later subgroup, moderator and meta-regression analysis.

Publication Bias

The next section examines the included studies for possible publication bias. This study used three approaches. The first is a funnel plot, which plots each study's effect size (k = 85) against its standard error.

Figure 3

Funnel Plot of Effect Sizes

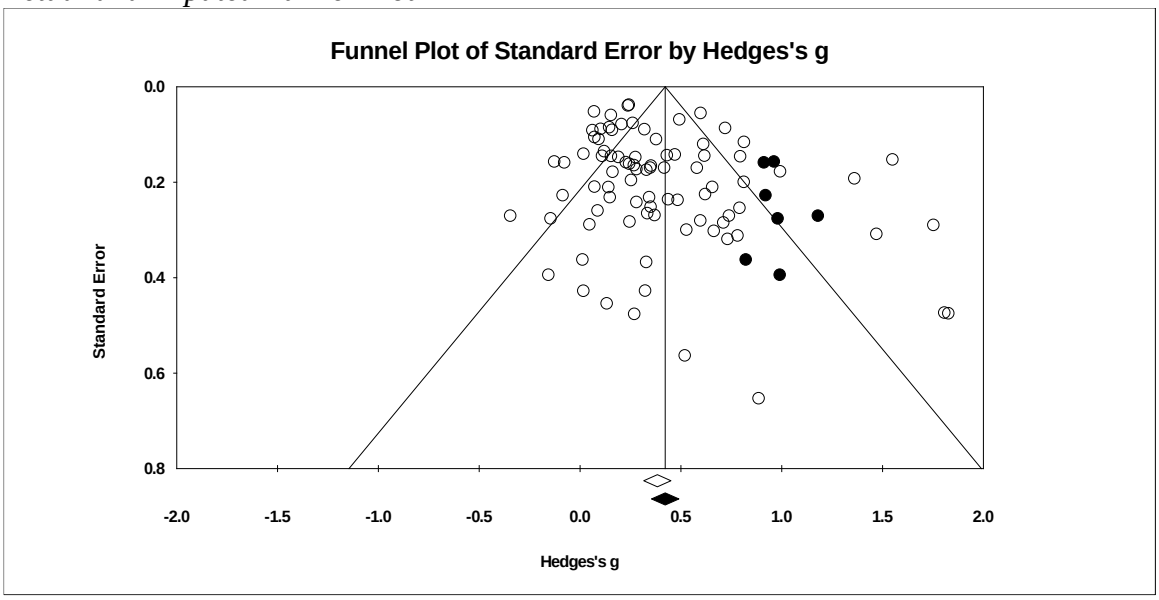


The funnel plot in figure 3 provides evidence of some asymmetry as a handful of studies outside the funnel towards the upper right-hand corner. These are the same studies

identified in the previous section. A second approach is by Duval and Tweedie (2000) is called the Trim and Fill methods and it uses the existing funnel plot and adds in any studies that the software would expect to see if. Results of the trim and fill test presents some evidence of publication bias in expecting an additional 7 studies in a more symmetric plot which would increase the mean study effect size from g = 0.38, 95% CL (0.32, 0.45) to g = 0.42, 95% CL (0.35, 0.49). These studies are represented as dark or filled in dots in the funnel plot in figure 4. The trim and fill approach suggests that given the plot of included studies that 'should' have been additional studies clustered towards the upper right, that would have increased the overall effect size. Thus the current dissertation may be taking a more conservative study of potential effect sizes of study abroad and intercultural competence.

Figure 4

Actual and Imputed Funnel Plot



Two related additional approaches to publication bias are the classic Fail-Safe *N* (Rosenthal, 1979) and Orwin's fail-safe *N* (1983). The Rosenthal approach focuses on the

number of studies that would be required to change the study *p*-value to 0. Using the Rosenthal approach, an 524 additional studies would be required to have an overall non-significant effect. A more conservative approach, Orwin's fail-safe *N* estimates how many studies with an effect size of '0' would be required to reduce the mean study effect size to what would be considered trivial (in this case using $g = 0.20$). This formula estimates that the next 49 studies would all have to have an effect size of 0 to reduce the overall effect size to $g = 0.20$.

Reviewing the analysis in the first research question the random-effects approach appears justified and significant heterogeneity was found. The overall mean study effect size was $g = 0.38$, a small to medium effect size and that there was limited evidence of publication bias.

Research Question 2: To what extent do individual student, study abroad program and research design characteristics moderate the relationship between study abroad and intercultural competence?

Research question two examines individual variables that might moderate the relationship of interest. Using Astin's IEO model, three groups of variables inputs (primarily student characteristics), environmental (characteristics of the study abroad program itself) and research design characteristics were studied. Each set was examined separately to gain an understanding of the specific relationship between the mean study effect size and that independent variable. This is an approach taken in almost all reviews included in the Campbell Collaboration the world-wide leader in systematic reviews and meta-analysis in social and behavioral sciences. These findings informed the variables to include in the meta-regression models in research question three.

This study used two types of analysis. The first is a meta-regression model, which is used for student variables like gender, program of study, and study duration. The meta-regression approach is used with continuous moderators. This approach indicates what impact an additional unit (1, or 10 or 100% of a student characteristic) or 1 additional week of study would have on a mean effect size. The second type of analysis is a traditional ANOVA model. This is used for most of the study abroad programs and research design characteristics where a category, such as the language used in the study, has multiple levels, but each study has only one level. Each study is only in a primary language, mostly in a primary language or primarily a foreign language. Whether the analysis is done with an ANOVA or regression technique, a random-effects model is always used. While each ANOVA or meta-regression does include a test of statistical significance, this study argues against a singular focus on *p*-values and more towards a comparison of the magnitude of effect sizes.

In this study, I focus primarily on effect sizes from tests using continuous moderators or the difference between categorical moderators where $g > 0.10$. This approach is used for five reasons. First, while I have done extensive work to identify as many potential studies as possible, there are only 85 effect sizes and detecting statistical significance with that relatively small sample is challenging. To that end, this study aims to provide support and recommendations for practitioners, policy makers and researchers and focussing solely on p-values above and below 0.05 does little to support those goals. In fact, it might present greater certainty than the evidence warrants. Secondly, many social science researchers use Cohen's (1998) heuristic of small (0.2 to 0.5), medium (0.5 to 0.8), large (0.8 and above) as guidelines for interpretations of effect sizes. This is

despite Cohen's warning that "there is a certain risk inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science" (1998, p. 25) and instead urged researchers to interpret effect sizes in the context of their own data. Thirdly, the National Survey of Student Engagement survey recommends institutions use a slightly different scales of small (0.1 to 0.3), medium (0.3 to 0.5) and larger (0.5 and above) provided that the estimate has significance at the 0.01 level (Rocconi & Gonyea, 2015). Moreover, effect sizes are most useful in comparison and a focus only on significances does not permit a comparison of relative differences. Again an effect size of 0.1 or more represents 1/3rd of the difference between a small and medium effect size using both heuristics. I argue that this is a difference worth exploring or noting. Finally, while some of these differences may or may not be significant, in each case all effect size estimates are accompanied by upper and lower 95% confidence limits, provided for transparency.

Testing moderators-Student characteristics.

Gender. The first variable is gender. In this example, the variable studied is the percentage of students identified as female within a study. Thus, the intercept would be the mean effect size of a study with only students who identify as male participating.

Table 12

Gender Moderator Effects

Variables (k = 85)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.63	-0.19	0.27, 1.00	3.38	<.01
Proportion Female	-0.37	0.27	-0.90, -0.16	-1.35	.18

Table twelve shows that the proportion of females does not significantly moderate the mean study effect size study as $Q(1) = 1.84$, $p = 0.18$, $\tau^2 = .06$, $I^2 = 78.88\%$, $R^2 = 0.00$. However, the finding may have practical significant later as if a study were 50% female

this would suggest an effect size about $g = 0.18$ smaller than a study of only male students.

Program of study. The second set of variables relates to the percentage of students in each of four areas of study. In the analysis below the intercept is the estimated effect size of a program with 100% social science students (the reference category).

Table 13

Program of Study Moderator Effects

Variables ($k = 85$)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.30	0.07	0.17, 0.43	4.36	< .001
Proportion Business	0.18	0.12	-0.06, 0.42	1.43	.15
Proportion STEM	0.17	0.15	-0.12, 0.47	1.16	.24
Proportion Other programs	0.09	0.10	-0.11, 0.29	0.89	.37

Results of the meta-regression for program of study, shows no moderating effects depending on the program of study and the overall model was not significant as $Q(3) = 2.65$, $p = 0.45$, $\tau^2 = .062$, $I^2 = 78.88\%$, $R^2 = 0.00$. There is a notable difference in effect sizes between proportion of business and STEM students compared to social science students.

Year of study. Each of proportion of students in first year, percentage of students in second year, percentage of students in fourth year and above, percentage not known, and percentage graduate students were put into a meta-regression model. In Table 14, the reference category is the percentage of 3rd year students in a study, meaning the intercept is interpreted as the estimated effect size of a study abroad program with 100% 3rd year students.

Table 14

Year of Study Moderator Effects

Variables (k = 85)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.32	0.08	0.16, 0.48	3.97	< .001
Proportion First year	-0.15	0.48	-1.10, 0.80	-0.31	.75
Proportion Second year	-0.01	0.17	-0.35, 0.33	-0.06	.95
Proportion Fourth+ year	0.13	0.18	-0.22, 0.48	0.72	.47
Proportion Unknown year	0.29	0.13	0.04, 0.54	2.25	.02
Proportion Graduate student	-0.22	0.42	-1.05, 0.61	-0.52	.61

The overall model test was not significantly associated with the mean study effect size $Q(5) = 7.26$, $p = 0.20$, $\tau^2 = .070$, $I^2 = 79.29\%$, $R^2 = 0.00$. This is not a meaningful change and indicates that the composition of students by year or level of study, does not moderate the mean effect size.

Ethnocultural identity. For ethnocultural identity, five categories of identity were coded. In Table 15, the proportion of students whose identity was not known is the reference category, meaning the intercept is the estimated effect size of a program with 100% students from an unknown ethnocultural identity.

Table 15

Ethnocultural/racial Identity Moderator Effects

Variables (k = 85)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.39	0.08	0.28, 0.50	7.20	< .001
Proportion Asian & Pacific Islander	0.23	0.23	0.67, 0.67	1.00	.32
Proportion African-American	-0.59	0.71	-1.98, 0.79	-0.84	.40
Proportion Hispanic/Latinx	-0.32	0.20	-0.72, 0.08	-1.57	.12
Proportion White	0.04	0.09	-0.14, 0.22	0.43	.67

The model did not find that the composition of students ethnocultural identity was associated with the mean study effect size $Q(4) = 4.29$, $p = 0.37$, $\tau^2 = .066$, $I^2 = 79.66\%$, $R^2 = 0.00$. This suggests that ethnocultural identity does not moderate the mean study effect size.

Domestic & international students. The next variables were the percentage of participants who were domestic or international students. For this variable only 56 studies provided enough data to do analysis. In this analysis, the intercept is what the study effect size would be if the study was of 100% international students.

Table 16

Student Status of Study Moderator Effects

Variables (k = 56)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.31	0.22	-0.12, 0.74	1.41	0.16
Proportion Domestic	0.13	0.23	-0.32, 0.58	0.58	0.56

The model above highlights that there was no significant association between the proportion of domestic or international students in a study and the mean study effect size $Q(1) = 0.34$, $p = 0.56$. $\tau^2 = .081$, $I^2 = 82.15\%$, $R^2 = 0.00$. However, a study with exclusively domestic students would be expected to have an effect size $g = 0.13$.

Previous experience abroad. The final variable in the student characteristics section are related to the percentage of students in a study who had previous work, study or travel experience abroad. Note that these estimates are also based on only 57 studies. In this analysis, the intercept is the estimated effect size if a study had no students with previous experience abroad.

Table 17

Previous Experience Abroad Moderator Effects

Variables (k = 57)	Coefficient	SE	CI	*Z*	*p*
Intercept	0.29	0.06	0.16, 0.42	4.50	< .001
Proportion Previously abroad	0.04	0.11	-0.18, 0.27	0.38	.70

The model above highlights that there was no significant association between the proportion of students in a study with previous experience abroad and the mean study effect size as $Q(1) = 0.14$, $p = 0.70$, $\tau^2 = .031$, $I^2 = 66.69\%$, $R^2 = 0.00$. This suggests that

the proportion of students who have previous experience abroad does not moderate the relationship.

Summarizing the findings in this section, no student variables were significantly associated with the mean effect size but there is evidence of a negative relationship for proportion of female students and proportion of social science students. The negative coefficients for ethnocultural identity (African-American and Hispanic/Latinx) likely warrants additional examination, though these findings should be viewed recalling the very large standard errors and remembering that about 5% students identified in both categories. But, overall there is not strong evidence that the student characteristics of the students in a study abroad program moderate the study effect size.

Testing moderators-Study abroad program characteristics. The next step is to investigate, primarily using ANOVA models, the association between the sub categories of the program characteristics. ANOVA models are primarily used in this section as most variables, except for program duration, are categorical.

Duration of study abroad program. Duration of study abroad program is calculated in a regression model as the variable is continuous and calculated in weeks. A random-effects meta-regression showed no significant association between study abroad duration and mean study effect size as $Q(1)= 0.06$, $p = .80$, $\tau^2 = .064$, $I^2 = 79.42\%$, $R^2 = 0.00$.

Table 18

Duration of Study Abroad Moderator Analysis

Variables ($k = 85$)	Coefficient	SE	CI	*Z*	*p*
Intercept	0.37	0.07	0.24, 0.50	5.57	< .001
Program duration (weeks)	0.00	0.01	-0.01, 0.01	0.25	.80

The rest of the study abroad program characteristics are analyzed using ANOVA models since each study effect size is categorized into only 1 sub-category of each characteristic (one type of language instruction, one type of faculty, one type of housing). This means that each sub-category has a result and there are no reference categories unlike the previous meta-regression models.

Language. The next variable is language of instruction. In this analysis studies were grouped as all English (k = 42) mostly English (k = 29) and primarily in an additional language (k = 14). While the test did not produce a significant result $Q(2)$= 3.79, p =. 15, studies not exclusive in a students primary language were associated with effect sizes between $0.13 < g < 0.15$ larger than those in their primary (usually English) language.

Table 19

Moderator Analysis of Language of Instruction

Variables (k = 85)	k	g	SE	CL	Z	p	$Q_b(df)$	p
Language of Instruction							3.79(2)	.15
Primary language	42	0.31	0.05	0.21, 0.41	6.06	<.001		
Mostly primary language	29	0.44	0.05	0.33, 0.55	8.06	<.001		
Primarily foreign language	14	0.46	0.09	0.28, 0.65	4.90	<.001		

Faculty instruction. The second set of variables looked at the faculty that taught the students. Among programs taught by home professors, local professors or a combination of the two, there may be some evidence of a relationship with the mean study effect size $Q(2)$= 4.50, p =. 11. However, we see that programs taught by home professors have a mean effect size about 50% larger than ones taught in combination.

Table 20

Moderator Analysis of Faculty Instruction

Variables (k = 85)	k	g	CL	Z	p	$Q_b(df)$	p
Faculty Instruction						4.50(2)	.11
Taught by home professors	33	0.47	0.36, 0.57	8.66	<.001		
Taught by a combination	40	0.31	0.21, 0.41	6.22	<.001		
Taught by local professors	12	0.39	0.19, 0.58	3.93	<.001		

Student housing arrangements. The third set of variables look at moderating impacts of housing arrangement. The options coded for housing where it was a combination or not known, a homestay, accommodation with other study abroad students and finally local students. Moderator analysis did not find any significant differences among the effect sizes of these groups $Q(3) = 0.07$ $p = .99$.

Table 21

Moderator Analysis for Housing Arrangements

Variables (k = 85)	k	g	CL	Z	p	$Q_b(df)$	p
Housing Arrangements						0.07(3)	.99
Combination/Not known	33	0.38	0.27, 0.49	6.94	<.001		
Homestay	18	0.40	0.23, 0.57	4.70	<.001		
W/ study abroad students	26	0.38	0.25, 0.50	5.70	<.001		
W/ local students	8	0.39	0.19, 0.60	3.72	<.001		

Nature of study abroad program. The fifth variable of interest was the nature of the program that students were participating in. One type would have a single group of students studying abroad together as a study (a faculty led program), the other being individual or several small groups in the same study (exchanges and consortiums). In this analysis, there was no relationship between the nature of the study abroad program and the mean effect size of the study as $Q(1) = 0.02$, $p = .90$.

Table 22

Moderator Analysis for Nature of Study Abroad Program

Variables (k = 85)	k	g	CL	Z	p	$Q_b(df)$	p
Nature of Study Abroad Program						0.02(1)	0.90

Individual/small groups	29	0.38	0.27, 0.49	6.78	<.001
Single group or large groups	56	0.39	0.30, 0.48	8.61	<.001

Intentional cultural interaction. The sixth variable looks at intentional cultural interaction in a study. In this example, intentional cultural interaction was not significantly associated with the mean study effect size $Q(1) = 0.46$, $p = .50$.

Table 23

Moderator Analysis for Intentional Cultural Interaction

Variables ($k = 85$)	k	g	CL	Z	p	$Q_b(df)$	p
Cultural Interaction						0.46(1)	0.50
Yes	57	0.40	0.32, 0.49	9.06	<.001		
No	28	0.35	0.24, 0.47	6.17	<.001		

Guided reflection. The last variable in this section is guided reflection, wherein staff or faculty part of the program intentionally engage in activities designed to help students reflect, process and learn from their experience abroad. Studies that had guided reflection do not have significantly different effect sizes from those that did not $Q(1) = .33$, $p = 0.56$.

Table 24

Moderator Analysis for Guided Reflection

Variables ($k = 85$)	k	g	CL	Z	p	$Q_b(df)$	p
Guided Reflection						0.33(1)	0.56
Yes	48	0.37	0.28, 0.46	7.90	<.001		
No	37	0.41	0.30, 0.51	7.71	<.001		

The section above conducted moderator analyses on seven characteristics of the study abroad program. While no overall models in this section about program characteristics were significant, a few areas suggest promising areas for additional investigation. Studies with the language mostly in the primary language or primarily in a foreign language showed larger effect sizes than those all in a student's primary

language. In addition, studies were the faculty were either from the home or local institution, showed larger effect sizes than those with a combination of faculty.

Testing moderators-Research design characteristics. The last section of examines research design as moderating characteristics. These are decisions and determinations by the researchers of a study about how they wanted to examine the phenomenon of study abroad and intercultural competence and can be viewed as potentially affecting internal and external validity. All these variables were analysed using ANOVA models.

Publication type. The first variable is publication type. Recall that conference papers were merged with journal articles, and that theses are merged with dissertations. The rationale for this is that theses and dissertations are developed by individual graduate students, who may become, but are not yet, faculty members and do not have the same publishing experience. Theses and dissertations are also more likely to be completed regardless of statistically significant findings; whereas conference papers and articles without significant results may not be submitted or accepted. In addition, conference papers, like journal articles are morel likely to subject to a blinded peer review process. Nevertheless, the analysis did not find a significant difference in mean study effect size by publication type as $Q(1) = 1.23$, $p = .27$, but this may provide limited evidence of publication bias as effect sizes for published articles are greater than dissertations.

Table 25

Moderator Analysis for Publication Type

Variables ($k = 85$)	K	g	CL	Z	p	$Q_b(df)$	p
Publication Type						1.23(1)	0.27
Article	41	0.42	0.33, 0.52	8.36	<.001		
Dissertation	44	0.35	0.25, 0.44	7.35	<.001		

Faculty involvement in research. The second variable is whether the author of the study (researcher) was also a participant in the study abroad program. The analysis did not find any significant difference by researcher involvement $Q(1) = 0.55$, $p = 0.46$.

Table 26

Moderator Analysis for Faculty involvement in Research

Variables (k = 85)	k	g	CL	Z	p	$Q_b(df)$	p
Faculty Involvement in research						0.55(1)	0.46
Not involved in program	47	0.36	0.28, 0.45	8.16	<.001		
Was involved res	38	0.42	0.31, 0.53	7.46	<.001		

Research instrument. The third analysis looks at moderating effects of the survey instrument chosen in a study. In this analysis, four specific instruments, plus 'others' were examined. Note the total number of effect sizes in this analysis is 89, not 85 like in previous analysis. This is because four studies used 2 instruments. The research instrument was not found to have a significant relationship to effect size $Q(4) = 7.91$, $p = .09$. However studies using the GPI and the IDI had mean effect sizes considerably ($g > 0.20$) lower than the CCAI, GMA and other instruments.

Table 27

Moderator Analysis for Research Instrument

Variables (k = 89)	k	g	CL	Z	p	$Q_b(df)$	p
Research Instrument						7.91	.09
GPI	6	.24	0.03, 0.45	2.28	.02		
IDI	42	.33	0.23, 0.43	6.36	<.01		
CCAI	9	.59	0.38, 0.80	5.46	<.01		
GMS	3	.61	0.21, 1.00	3.00	<.01		
Other	29	.41	0.30, 0.52	7.33	<.01		

Research design. The final moderator analysis focuses on the research design. The three options were retrospective, meaning that the pre-test was given at the end of the study abroad program at the same time as the post-test, traditional pre-test/post-test and

pre-test/post-test with a comparison group. The analysis shows that research design has a significant relationship to variation in mean study effect size $Q(2) = 18.23$, $p < .01$. Retrospective pre-test design ($k = 2$) had higher effect sizes, than pre-post designs ($k = 62$) or pre-post with control groups ($k = 21$). While only two studies used this approach (Armfield, 2004; Palmer, 2009) it may serve as a caution for consumers of research or future researchers of the impact this decision may have on the estimates or outcomes of a study. The difference of the mean effect size of studies that used a pre-test/post-test approach versus a pre-test/post-test comparison approach was trivial.

Table 28

Moderator for Research design

Variables	*k*	*g*	CL	Z	*p*	$Q_b(df)$	*p*
Research design						18.23	<.01
Retrospective pre-test	2	1.17	0.80, 1.54	6.24	<.001		
Pre-test/post-test designs	62	0.37	0.29, 0.44	9.84	<.001		
Pre-test/post-test comparison designs	21	0.36	0.21, 0.51	4.75	<.001		

The moderators relating to research design resulted in several significant and meaningful relationships. A significant relationship was found for research instrument with studies that used the IDI and GPI associated with considerably lower effect sizes than studies that used the CCAI and GMS. Also, studies that used retrospective pre-tests were associated with significantly larger effect sizes than studies that use traditional pre-test/post-test or pre-test/post-test with comparison group designs.

Summary of Moderator Effects

In the previous sections, three sets of moderators (student, study abroad program and research design characteristics) were tested independently. The results showed that few variables individually help to understand variation in mean study effect sizes. Still,

there were examples in each section that highlighted some practical importance including gender and program of study (student characteristics), language of instruction and faculty instruction (study abroad program characteristics) and survey instrument and research design (research design characteristics) The next step is to use sets of variables together in meta-regression models. This allows for testing multiple variables at the same time and for analyzing categorical and continuous variables sequentially in the same model. The meta-analytic approach, like multiple regression, permits testing of potentially confounding predictor variables.

Research Question 3: To what degree do combinations of student, study abroad program and research design characteristics, and all characteristics explain variance in the mean study effect size?

The third research questions consists of four sets of meta-regressions. This approach allows multiple variables in a model simultaneously so that they are not considered in isolation as in the second research question. The first analysis will focus on the variables from the student characteristics section.

Meta regression of student characteristics. In previous analysis, the relationship of gender, program of study, year of study, ethnocultural identity, domestic and international status, and previous experience abroad to study effect sizes were examined. In total there were 21 individual variables that were coded for this section. Before beginning the meta-regression a correlation table of the student characteristics and the student effect sizes were plotted as shown below. The correlation matrix included all variables in the section and used a very liberal alpha of $p = .1$ to denote significance.

Table 29

Correlation matrix of Student Characteristics

Variables ($k = 85$)	g	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2) Proportion Female	-0.05							
(3) Proportion Business	0.08	-0.34*						
(4) Proportion Social Science	-0.19*	0.26*	-0.36*					
(5) Proportion STEM	0.15	0.02	-0.18	-0.29*				
(6) Proportion Other program	0.02	-0.00	-0.34*	-0.54*	-0.23*			
(7) Proportion 1st Year	-0.10	0.09	-0.05	0.06	-0.03	0.00		
(8) Proportion 2nd Year	-0.12	0.08	-0.04	0.22*	-0.10	-0.13	0.26*	
(9) Proportion 3rd Year	-0.18	-0.29*	0.10	-0.30*	-0.15	0.34*	-0.19*	-0.26*
(10) Proportion 4th Year	0.05	0.17	0.20*	0.05	0.25*	-0.38*	-0.08	-0.25*
(11) Proportion Other Year	0.24*	-0.04	-0.17	-0.01	0.06	0.11	-0.15	-0.26*
(12) Proportion Graduate	-0.01	0.29*	-0.09	0.24*	-0.09	-0.11	-0.03	-0.13
(13) Proportion Domestic	0.06	0.13	-0.25*	0.11	0.04	0.08	0.11	0.10
(14) Proportion Asian & Pacific Islander	0.12	0.26*	-0.07	0.10	-0.06	-0.00	0.25*	0.34*
(15) Proportion African-American	-0.06	0.08	-0.16	0.06	0.23*	-0.09	0.20*	0.08
(16) Proportion Hispanic/Latinx	-0.13	-0.02	-0.12	-0.02	0.29*	-0.07	-0.07	-0.10
(17) Proportion White	0.07	-0.09	-0.00	-0.00	0.17	-0.11	0.01	0.10
(18) Proportion Unknown Identity	-0.06	-0.02	0.10	-0.04	-0.26*	0.14	-0.10	-0.20*
(19) Proportion Previously Abroad	0.06	-0.12	0.03	-0.22*	0.12	0.15	0.00	0.00

* $p < .1$

Table 29

Correlation matrix of Student Characteristics continued

Variables ($k = 85$)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(10) Proportion 4th Year	-0.18*									
(11) Proportion Other Year	-0.50*	-0.39*								
(12) Proportion Graduate	-0.23*	-0.02	-0.05							
(13) Proportion Domestic	0.03	0.06	-0.12	-0.10						
(14) Proportion Asian & Pacific Islander	-0.24*	-0.17	0.08	-0.04	0.10					
(15) Proportion African-American	-0.16	0.17	-0.07	-0.01	0.09	-0.01				
(16) Proportion Hispanic/Latinx	0.14	-0.09	-0.09	0.30*	0.04	-0.05	-0.05			
(17) Proportion White	-0.12	0.15	-0.08	0.03	0.22	-0.13	0.26*	-0.11		
(18) Proportion Unknown Identity	0.17	-0.05	0.08	-0.12	-0.28*	-0.26*	-0.32*	-0.26*	-0.84*	
(19) Proportion Previously Abroad	0.12	-0.01	0.04	-0.29*	-0.10	0.11	-0.05	-0.20	0.03	0.01

* $p < .1$

Table 29 shows a negative correlation between the percentage of students in a social science program and the mean study effect size. The analysis found a positive relationship between the proportion of students in "other" years of study and the mean effect size. Reviewing the rest of the table did not highlight any other cases where the correlation between potential moderators was large. Based on the earlier meta-regression models and the above correlation table, three variables were chosen for the student characteristics meta-regression model. They are proportion of female students, proportion of students in social science programs and proportion of students in third year. This last variable was chosen as a proxy for a midpoint in year of study and because it was the largest group. No ethnocultural variables were included in these meta-regression models as nearly half of the participants in the study did not have an identified ethnocultural identity. In this meta-regression model, the intercept would be a study of 100% male, 100% in social science and 100% 3rd year of study.

Table 30

Meta-Regression for Student Characteristics

Variables (k = 85)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.32	0.26	-0.18, 0.83	1.25	.21
Gender					
Proportion Female	-0.36	0.35	-0.85, 0.51	-0.50	.62
Program of study					
Proportion Business students	0.19	0.15	-0.10, 0.48	-1.28	.20
Proportion Social Science Students			Reference Group		
Proportion STEM students	0.14	0.17	-0.18, 0.47	0.85	.40
Proportion Other programs	0.10	0.12	-0.14, 0.33	0.79	.43
Year of Study					
Proportion 1st year students	-0.07	0.51	-1.07, 0.92	-0.14	0.89
Proportion 2nd year students	0.08	0.20	-0.31, 0.47	0.41	0.68
Proportion 3rd year students			Reference Group		
Proportion 4th year students	0.15	0.22	-0.27, 0.57	0.71	0.48
Proportion Other years	0.33	0.14	0.06, 0.60	2.39	0.02
Proportion Graduate students	-0.01	0.45	-0.90, 0.87	-0.03	0.98

The results of the meta-regression suggest that the model and the included covariates do not help to explain the variance in mean effect sizes of the included studies $Q(9) = 9.87$, $p = 0.36$ and that the model fit test provided estimates of $Q(75) = 365.59$, $\tau^2 = .078$, $I^2 = 79.49\%$, $R^2 = 0.00$. While the overall model was not significant, the estimates show a large negative coefficient for the proportion of female students in a study suggesting studies with more male students were associated with larger effect sizes. In additional, for program of study, all programs of study have positive coefficients suggesting, studies with larger proportions of social science students have smaller effect sizes. Finally, while there was a significant result for proportion of other years compared to 3rd year students, the coefficient was very small, and no other year of study variables had similar size estimates, and all had large *p*-values.

Meta regression on study abroad program characteristics. The second meta-regression focuses on study abroad program (environment) characteristics of the study. Recall that no moderator categories individually were significantly associated with mean effect sizes of the study. The study abroad program characteristics were put into a correlation matrix with the results in Table 31.

Table 31

Correlation Matrix of Study Abroad Program Characteristics

Variables (k = 85)	g	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2) Duration (weeks)	0.01							
(3) Primary language instruction	-0.16	-0.20*						
(4) Mostly primary language	0.12	0.13	-0.72*					
(5) Primarily foreign language	0.06	0.11	-0.45*	-0.30*				
(6) Home faculty	0.24*	0.25*	-0.19*	0.32*	-0.16			
(7) Combination faculty	-0.23*	-0.20*	0.04	-0.11	0.10	-0.74*		
(8) Local faculty	-0.02	-0.06	0.21*	-0.29*	0.08	-0.33*	-0.38*	
(9) Combo/Unknown housing	-0.00	-0.01	0.13	0.03	-0.22*	0.15	-0.05	-0.14
(10) Homestay	-0.02	0.14	-0.21*	-0.13	0.45*	-0.03	-0.04	0.10
(11) With study abroad students	0.02	-0.23*	0.20*	0.01	-0.29*	-0.23*	0.17	0.07
(12) With local students	0.00	0.20*	-0.25*	0.12	0.19*	0.15	-0.13	-0.02
(13) Single or large groups	0.02	-0.07	0.21*	-0.38*	0.20*	-0.26*	0.13	0.18*
(14) Cultural interaction	0.10	-0.08	0.00	-0.15	0.19*	0.01	0.06	-0.10
(15) Guided reflection	-0.07	-0.06	-0.15	0.00	0.20*	0.03	0.06	-0.14

* $p < .1$

Correlation Matrix of Study Abroad Program Characteristics continued

Variables (k = 85)	(9)	(10)	(11)	(12)	(13)	(14)
(10) Homestay	-0.42*					
(11) With study abroad students	-0.53*	-0.34*				
(12) With local students	-0.25*	-0.16	-0.21*			
(13) Single or large groups	-0.31*	0.39*	-0.01	-0.01		
(14) Cultural interaction	-0.33*	0.38*	0.02	-0.02	0.39*	
(15) Guided reflection	-0.15	0.42*	-0.24*	0.05	0.12	0.67*

* $p < .1$

In the correlation matrix, the type of faculty instruction was significantly correlated with the study outcome. Language of instruction was not significantly

correlated but the moderator estimates did show practically significant differences in effect sizes and these variables will be includes. Finally, neither cultural interaction, nor guided reflection variables were significant, but they are strongly linked to the Allport theory and along with instruction and language will be used in the meta-regression model on program characteristics. In the model, found in table 32, the reference categories are study abroad programs in a students primary language (usually English), programs taught by home faculty and programs with no cultural interaction and no guided reflection.

Table 32

Meta-Regression Summary for Program Characteristics

Variables (k = 85)	*Coefficient*	SE	CI	*Z*	*p*
Intercept	0.38	0.10	0.18, 0.57	3.84	<0.01
Language of Instruction					
Primary language			Reference	Group	
Mostly primary language	0.12	0.09	-0.05, 0.29	1.33	0.18
Primarily foreign language	0.17	0.11	-0.05, 0.40	1.54	0.12
Faculty Instruction					
Home faculty			Reference	Group	
Combination faculty	-0.15	0.08	-0.31, 0.02	-0.77	0.44
Local faculty	-0.07	0.13	-0.33, 0.19	-0.55	0.58
Cultural Interaction					
Yes	0.18	0.10	-0.02, 0.38	1.75	0.08
Guided Reflection					
Yes	-0.17	0.10	-0.37, 0.03	-1.70	0.09

The results of the meta-regression suggest that the model does not significantly help to explain the variance in mean effect sizes of the included studies $Q(6) = 10.31$, $p = 0.11$ and that the model fit test provided estimates of $Q(78) = 372.48$, $\tau^2 = .070$, $I^2 = 79.06\%$, $R^2 = 0.00$. While the overall model was not significant, the estimates show positive coefficients for programs of study that had some element of foreign language instruction. Secondly, programs taught only by a combination of home and host faculty had lower effect sizes than studies of home or local faculty. Finally, studies that had cultural

interaction had larger effect sizes, but studies with guided reflection had small effect sizes.

Meta regression of research design characteristics. The third meta-regression focuses on research design characteristics and how they may be related to variation in mean study effect sizes. A correlation matrix in Table 33 of all the individual research design characteristics was used to identify possible relationships for inclusion in the meta-regression. Negative relationships were found for studies using the IDI instrument and a positive relationship were found for studies using the CCAI. Finally, a very large effect (positive) relationship was found for studies using retrospective pre-test designs. Thus, variables included in the meta-regression included survey instrument and research design, but also publication type as a check against publication bias (relating to guiding framework of validity theory). In the meta-regression in table 34, the reference categories were dissertations (publication type), IDI instrument (survey instrument) and pre-test/post-test designs (research designs).

Table 33

Correlation matrix of Research Design Characteristics

Variables ($k = 89$)	g	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(2) Article	0.13											
(3) Dissertation	-0.13	-1.00										
(4) Faculty Not Involved	-0.13	-0.80*	0.80*									
(5) Faculty Involved	0.13	0.80*	-0.80*	-1.00								
(6) GPI Instrument	-0.10	0.02	-0.02	0.15	-0.15							
(7) IDI instrument	-0.18*	0.10	-0.10	-0.25*	0.25*	-0.25*						
(8) CCAI Instrument	0.19*	0.06	-0.06	0.07	-0.07	-0.09	-0.32*					
(9) GMS Instrument	0.05	-0.05	0.05	0.04	-0.04	-0.05	-0.18*	-0.06				
(10) Other Instrument	0.10	-0.13	0.13	0.13	-0.13	-0.19*	-0.66*	-0.23*	-0.13			
(11) Retrospective pre-test	0.27*	-0.14	0.14	0.13	-0.13	-0.04	0.01	-0.05	-0.03	0.06		
(12) Pre-test/Post-test	-0.11	0.34*	-0.34*	-0.20*	0.20*	0.07	0.09	-0.04	-0.16	-0.05	-0.24*	
(13) Pre-test/Post-test Comparison	0.02	-0.30*	0.30*	0.16	-0.16	-0.06	-0.10	0.06	0.17	0.03	-0.09	-0.94*

* $p < .1$

Table 34

Meta-Regression Summary for Research Design Characteristics

Variables (k = 89)	Coefficient	SE	CI	Z	p
Intercept	0.21	0.06	0.09, 0.34	3.84	0.01
Publication Type					
Article	0.14	0.07	0.01, 0.27	2.13	0.03
Survey Instrument					
GPI Instrument	-0.04	0.11	-0.25, 0.17	-0.40	0.69
IDI Instrument			Reference	Group	
CCAI Instrument	0.28	0.11	0.07, 0.50	2.57	0.01
GMS Instrument	0.32	0.20	-0.06, 0.71	1.64	0.10
Other Instrument	0.10	0.07	-0.04, 0.24	1.46	0.14
Research Design					
Retrospective pre tests	0.89	0.19	0.53, 1.26	4.80	<0.01
Pre-test/post-test			Reference	Group	
Pre-test/post-test comparison	0.02	0.08	-0.14, 0.19	0.27	0.78

The results of the meta-regression for research design characteristics suggest that the model does significantly help to explain the variance in mean effect sizes of the included studies $Q(7) = 34.7$, $p = < 0.01$ and that the model fit test provided estimates of $Q(81) = 280.83$, $\tau^2 = .046$, $I^2 = 71.16\%$, $R^2 = 0.26$. A significant result was found for articles, meaning that adjust for other research design characteristics, they had significantly larger effect sizes than dissertations ($g = 0.14$). Secondly, survey instruments had significant results. Studies that used the CCAI or the GMS had significantly larger effect sizes than those that used the IDI, while studies that used the GPI had very similar findings. Finally, studies that used retrospective pre-tests had significantly larger effect sizes than studies that used either form of pre-test/post-test designs.

Meta regression of all sets of characteristics. A final step in this section was to run a model that included variables from all three sections. The first step is to construct a single correlation matrix of all potential variables. Significant relationships were found for language of instruction, survey instrument and research design. A model was

developed to focus on characteristics from all three sets of moderators. From the student characteristics gender (though not significant) and program of study are included. In the student characteristics section language of instruction (which was significantly correlated), cultural interaction and guided reflection were included. The latter two for theoretical reasons related to Allport's theory and Engle and Engle's framework. Finally instrument type and research design were included from the research design characteristics. In this model, the reference categories are proportion of male students, proportion of social science students, programs taught in a students primary language, studies using the IDI and studies with a pre-test/post-test design.

Table 35

Correlation matrix of All Study Characteristics

Variables ($k = 89$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(2) Proportion Female	-0.03										
(3) Proportion Business	0.09	-0.33*									
(4) Proportion Social Science	-0.17	0.29*	-0.35*								
(5) Proportion STEM	0.14	0.03	-0.18	-0.27*							
(6) Proportion other program	-0.00	-0.04	-0.35*	-0.55*	-0.23*						
(7) Primary language instruction	-0.16	0.00	0.20*	0.04	0.08	-0.25*					
(8) Mostly primary language	0.12	-0.08	-0.11	-0.03	0.02	0.11	-0.72*				
(9) Primarily foreign language	0.06	0.10	-0.13	-0.01	-0.13	0.20*	-0.45*	-0.30*			
(10) Home faculty	0.24*	-0.15	0.10	-0.04	-0.04	-0.02	-0.19*	0.32*	-0.16		
(11) Combination faculty	-0.23*	-0.08	-0.11	-0.13	0.03	0.20*	0.04	-0.11	0.10	-0.74*	
(12) Local faculty	-0.02	0.33*	0.02	0.24*	0.00	-0.26*	0.21*	-0.29*	0.08	-0.33*	-0.38*
(13) Cultural Interaction	0.10	0.19*	-0.07	-0.03	0.02	0.07	0.00	-0.15	0.19*	0.01	0.06
(14) Guided reflection	-0.07	0.24*	-0.11	0.03	-0.17	0.16	-0.15	0.00	0.20*	0.03	0.06
(15) Article	0.13	0.01	0.17	0.03	-0.13	-0.09	0.06	-0.18*	0.15	0.02	-0.02
(16) Dissertation	-0.13	-0.01	-0.17	-0.03	0.13	0.09	-0.06	0.18*	-0.15	-0.02	0.02
(17) GPI Instrument	-0.10	-0.08	0.19*	-0.15	0.09	-0.06	-0.10	0.10	0.01	0.06	-0.07
(18) IDI Instrument	-0.18*	0.07	-0.03	-0.02	-0.03	0.06	-0.03	-0.18*	0.27*	-0.02	0.12
(19) CCAI Instrument	0.19*	-0.04	0.09	0.15	-0.12	-0.15	0.03	0.01	-0.04	-0.04	-0.16
(20) GMS Instrument	0.05	-0.05	-0.10	0.24*	-0.05	-0.12	0.18*	-0.13	-0.08	-0.15	0.08
(21) Other Instrument	0.10	0.01	-0.09	-0.09	0.09	0.11	0.00	0.18*	-0.23*	0.08	-0.02
(22) Retrospective pre-test	0.27*	-0.03	-0.04	-0.05	-0.01	0.09	-0.01	0.06	-0.07	0.19*	-0.14
(23) Pre-test/post-test	-0.11	-0.01	0.17	-0.35*	-0.10	0.29*	-0.00	-0.05	0.06	-0.16	0.18*
(24) Pre-test/post-test comparison	0.02	0.02	-0.16	0.38*	0.11	-0.33*	0.01	0.03	-0.04	0.10	-0.13

* $p < .1$

Table 35

Correlation matrix of All Study Characteristics continued

Variables ($k = 89$)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
(13) Cultural Interaction	0.06	-0.10											
(14) Guided reflection	-0.06	-0.14	0.67*										
(15) Article	-0.02	-0.01	0.31*	0.18									
(16) Dissertation	0.02	0.01	-0.31*	-0.18	-1.00								
(17) GPI Instrument	-0.07	0.02	-0.09	-0.03	0.02	-0.02							
(18) IDI Instrument	0.12	-0.14	0.27*	0.27*	0.10	-0.10	-0.25*						
(19) CCAI Instrument	-0.16	0.28*	0.09	-0.07	0.06	-0.06	-0.09	-0.32*					
(20) GMS Instrument	0.08	0.10	-0.12	-0.08	-0.05	0.05	-0.05	-0.18*	-0.06				
(21) Other Instrument	-0.02	-0.08	-0.25*	-0.19*	-0.13	0.13	-0.19*	-0.66*	-0.23*	-0.13			
(22) Retrospective pre-test	-0.14	-0.06	-0.05	-0.02	-0.14	0.14	-0.04	0.01	-0.05	-0.03	0.06		
(23) Pre-test/post-test	0.18*	-0.02	0.17	0.24*	0.34*	-0.34*	0.07	0.09	-0.04	-0.16	-0.05	-0.24*	
(24) Pre-test/post-test comparison	-0.13	0.05	-0.16	-0.24*	-0.30*	0.30*	-0.06	-0.10	0.06	0.17	0.03	-0.09	-0.94*

* $p < .1$

Table 36

Meta-Regression of All Study Characteristics

Variables (k = 89)	Coefficient	SE	CI	Z	p
Intercept	-0.07	0.24	-0.54, 0.41	-0.28	0.78
Gender					
Proportion female	0.09	0.30	-0.49, 0.67	0.30	0.76
Program of Study					
Proportion Business	0.37	0.13	0.11, 0.67	2.79	<0.01
Proportion Social Science			Reference	Group	
Proportion STEM	0.33	0.15	0.04, 0.62	2.24	0.03
Proportion other program	0.11	0.10	-0.09, 0.31	1.06	0.29
Language of Instruction					
Primary language			Reference	Group	
Mostly primary language	0.21	0.08	0.06, 0.36	2.81	<0.01
Primarily foreign language	0.32	0.11	0.11, 0.53	2.95	<0.01
Cultural Interaction					
Yes	0.13	0.10	-0.06, 0.31	1.32	0.19
Guided Reflection					
Yes	-0.09	0.10	-0.40, 0.06	-0.88	0.38
Survey Instrument					
GPI Instrument	-0.04	0.10	-0.28, 0.10	-1.43	0.15
IDI Instrument			Reference	Group	
CCAI Instrument	0.28	0.12	0.07, 0.53	2.52	<0.01
GMS Instrument	0.32	0.22	0.18, 1.02	2.79	<0.01
Other Instrument	0.10	0.08	-0.06, 0.25	1.23	0.22
Research Design					
Retrospective pre-test	0.87	0.19	0.50, 1.24	4.55	<0.01
Pre-test/post-test			Reference	Group	
Pre-test/post-test comparison	-0.02	0.09	-0.20, 0.16	-0.21	0.84

The results of the meta-regression for a combined model of all sets of characteristics suggest that the model does significantly help to explain the variance in mean effect sizes of the included studies $Q(14) = 49.14$, $p = < 0.01$ and that the model fit test provided estimates of $Q(74) = 246.45$, $\tau^2 = .050$, $I^2 = 69.97\%$, $R^2 = 0.18$. Significant results were found for program of study in that both the proportion Business and STEM students were associated with larger effect sizes for social science students. In addition, studies where the students were taking some foreign language courses (mostly primary and primarily

foreign language) were associated with larger gains intercultural competence than studies in a student primary language. Thirdly, significant results were found for survey instrument with studies using the IDI scoring significantly smaller coefficients than those using the CCAI or GMS instruments. Finally, studies that used retrospective pre-tests had significantly larger effect sizes than studies that used pre-test/post-test designs. In the model above, gender was not a meaningful predictor, thus an additional model was run, removing gender and the estimates are found in table 37.

Table 37

Follow-up Meta-Regression of All Study Characteristics

Variables (k = 89)	Coefficient	SE	CI	Z	p
Intercept	-0.00	0.10	-0.20, 0.20	-0.00	0.99
Program of Study					
Proportion Business	0.35	0.12	0.12, 0.59	2.98	<.01
Proportion Social Science			Reference	Group	
Proportion STEM	0.32	0.14	0.04, 0.60	2.24	.03
Proportion other program	0.10	0.10	-0.09, 0.30	1.02	.31
Language of Instruction					
Primary language			Reference	Group	
Mostly primary language	0.21	0.07	0.06, 0.35	2.83	<0.01
Primarily foreign language	0.31	0.11	0.11, 0.52	2.96	<0.01
Cultural Interaction					
Yes	0.12	0.09	-0.06, 0.31	1.31	0.19
Guided Reflection					
Yes	-0.08	0.09	-0.26, 0.10	-0.84	0.40
Survey Instrument					
GPI Instrument	-0.17	0.12	-0.40, 0.06	-1.44	0.15
IDI Instrument			Reference	Group	
CCAI Instrument	0.30	0.12	0.07, 0.53	2.54	0.01
GMS Instrument	0.59	0.21	0.18, 1.00	2.82	<0.01
Other Instrument	0.10	0.08	-0.06, 0.25	1.23	0.22
Research Design					
Retrospective pre-test	0.87	0.19	0.50, 1.24	4.61	<0.01
Pre-test/post-test			Reference	Group	
Pre-test/post-test comparison	-0.02	0.09	-0.20, 0.16	-0.23	0.82

The second meta-regression of all sets of characteristics suggest that the model does significantly help to explain the variance in mean effect sizes of the included studies $Q(13) = 50.15$, $p = < 0.01$ and that the model fit test provided estimates of $Q(75) = 247.42$, $\tau^2 = .048$, $I^2 = 69.69\%$, $R^2 = 0.21$. This model is a small improvement over the one in table 34. However while the analog R^2 increased slightly, none of the coefficients changed direction, all estimates that were significant in the first model remain significant and the estimates for most variables are almost identical to the previous. Thus the results of the combined meta-analysis provide some significant different results than the results of the individual moderators in research question two and even the three sets of meta-regressions earlier in chapter. For example, program of study produced significant results in the above analysis, which did not appear earlier. This also true for language of instruction which is significant in the model above but not in the study abroad program characteristics meta-regressions. This highlights the importance of considering a set of variables that more broadly reflect the characteristics of a single study and analyzing them using an IEO framework. Examining the variables only in isolation or without consideration of how variables from the other two categories may be correlated with them can produce very different estimates.

Conclusion

In this chapter, 72 studies, provided 85 independent (plus four additional dependent) effect sizes for analysis. Descriptive statistics and frequencies were provided for all potential moderator variables. An overall random effects model was conducted to produce a summary effect size of $g = 0.38$, $df(84)$, 95% CI (0.32, 045), $p <.001$. In

addition, substantial heterogeneity was observed as $Q(84) = 404.75$, $p < .01$ supporting the decision to use a random-effects approach.

Using a stem and leaf plot, a few studies emerged as possible outliers but as no common characteristics among them could be identified they were retained. Funnel plots and trim and fill tests highlighted some evidence of possible publication bias with studies having effect sizes larger than the mean study effect size most likely to be missing. Fail safe N tests suggested it would require more than 500 studies to produce a non-significant overall estimate and nearly 50 studies with an effect size of zero to reduce the mean study effect size to $g = 0.20$. Three sets of moderator analyses were conducted using both meta-regression (primarily the student characteristics) and sub-group analysis approaches (program and research design characteristics), plus an additional omnibus set of models. Chapter five concluded with a series of meta-regressions, one for each set of characteristics. The final chapter will discuss these results as they relate to existing literature, including the 2017 Varela meta-analysis, the chosen guiding frameworks and conclude with implications for practice, policy and research.

Chapter Six: Discussion and Conclusion

This dissertation attempted to gain a better understanding of three research questions about the relationship between study abroad participation and intercultural competence. The final chapter reviews the results from the previous chapter, discuss how they intersect with existing literature, including the 2017 meta-analysis conducted by Varela. The limitations of the study will be detailed, before discussing the implications for study abroad programs, policy-makers and researchers, followed by concluding thoughts.

Research Question 1

To what degree do combinations of student, study abroad program and research design characteristics, and all characteristics explain variance in the mean study effect size?

The foundational question of the study sought to understand if a relationship exists between student abroad and intercultural competence, the strength of that relationship and if there is evidence of significant variation in the mean study effect sizes. The hypothesis was confirmed and overall, the results of this study suggest that there is a positive relationship between study abroad participation and the development of intercultural competence. The results of the random effects model analysis of 85 effect sizes was $g = 0.38$, which, using Cohen's heuristic, would be considered approximately halfway between a small and medium effect size. Furthermore, heterogeneity tests identified considerable variance between studies. With this result, additional analysis on the moderators identified in the literature review and from the IEO and Engle and Engle models followed.

The finding of a small and significant effect size supports most of the published research on the positive relationship between study abroad and intercultural competence

development (see Black & Bernardes, 2014; Gingerich, 1998; Palmer, 2009; Zarges, 2016) and challenges other studies that found no significant relationship between study abroad and intercultural competence like Burrow (2010), Keefe (2008) and Sell (1983). One factor that distinguishes the current study from much of the existing literature is the inclusion of an effect size that describes not just the direction of the relationship but the strength or magnitude of it. Few studies on study abroad and intercultural competence have gone beyond the simple test of significance. The meta-analysis done by Varela (2017), found an effect size of $d = 0.46$, slightly larger than found in this study. However, the Varela finding is based on only 30 studies and 38 effect sizes, compared to the 72 studies and 85 effect sizes in this dissertation. Contrasts between the current student and the Varela meta-analysis will be discussed in greater detail later.

Research Question 2

Research question two was comprised of three sections questions, one relating to each set of individual student, study abroad program, and research design characteristics.

Research Question 2) To what extent do individual student population characteristics moderate the relationship between study abroad and intercultural competence?

The six student level moderators were gender, program of study, year/level of study, ethnocultural/racial identity, international/domestic status and previous experience abroad. The null hypothesis was that no student characteristics would be significantly associated with variance in effect sizes. Using basic moderator tests four variables (year of study, domestic/international students, ethno-cultural/racial identity and previous experience abroad) supported the hypothesis of no meaningful relationship. However, two variables, gender and program of study, while not producing statistically significant results had practically significant effect sizes. A study of all female students would be

expected to have an effect size $g = 0.37$ lower than a study with all male students. The finding relating to gender challenges earlier work by Nichols (2011), Rexeisen and Al-Khatib (2008), and Zarges (2016). These three studies all found larger effects for female students than male students. Secondly, a study where all of the students were in business (of which there were several in this dissertation) or STEM programs associated with a mean effect size of $g = 0.17$ and 0.16 larger, respectively, than a study of all social science students. The finding for program of study challenges previous work from Vande Berg (2009) of a negative association for business students, as the current study presents a positive result. However, in both cases, the findings in the student characteristics section are univariate, not multivariate, results and it is possible that other variables may confound these findings.

Research Question 2) To what extent do individual study abroad program characteristics moderate the relationship between study abroad and intercultural competence?

The second part of this research question sought to examine which, if any, study abroad program characteristics were associated with mean effect size. The hypothesis was that all variables, except for duration, would be associated with larger effect size, as they moved closer to level 5 in the Engle and Engle framework. Duration, measured in weeks, was not significantly associated with variance in the mean effect sizes; supporting the original hypothesis of no relationship. Nor was type of housing, presence of in cultural interaction or guided reflection. Two sets of variables, language used in the study abroad program and nature of the faculty instruction, had promising results. Studies that had at least some foreign language instruction were associated with larger effect sizes than programs entirely in a student's primary language (typically English). Secondly, students in programs led or taught by faculty from their home institution or local faculty

had larger effect sizes than studies with mixed faculty. Neither were significant; however both produced a meaningful effect size difference. Overall this suggests qualified support for the Engle and Engle classification system. There is no evidence that program characteristics in levels 1 and 2 are associated with lower effect sizes. But there is some evidence that supports the hypothesis that higher levels in the areas of language, faculty who teach the program are associated with larger effect sizes. This also suggests qualified support for the Intergroup Contact Theory.

Research Question 2) To what extent do individual research design characteristics moderate the relationship between study abroad and intercultural competence?

The last set of moderators examined were research design characteristics. The initial hypothesis was that survey instrument, and research design would be significantly associated with mean effect sizes. Overall, no significant results were found for publication type (though articles did have larger effect sizes than dissertations), nor for faculty involvement in the research study (but when faculty were involved, effect sizes were higher). There was no significant result for research instrument, but studies using the IDI scored lower than the CCAI, the GMS and all other instruments (except the GPI). A significant result was found that studies using a retrospective pre-test design scored significantly higher than those using a pre-test/post-test design and a pre-test/post-test comparison design.

Thus the hypothesis for larger effect sizes for dissertations is not supported (possible weak evidence of publication bias) but there may be weak support that researcher involvement in the study abroad program is associated with larger effect sizes. The hypothesis about the IDI was largely supported though one instrument, the GPI had even average effect sizes lower. And the hypothesis for research design was partially

supported as retrospective studies scored significantly higher but there was no meaningful difference between pre-post designs and pre-post-comparison designs.

Overall, across all three levels of moderators, only a few individual moderators were significantly associated with mean study effect size. But there are several cases where the coefficient or effect size would be interpreted as meaningful; where differences of $g > 0.10$ or more was found. However, the practice of focussing on individual moderators ignores the multi-faceted identities of the participants in a program, the nuances of study abroad program design and the choices made in developing a study by the researcher. This is one of the primary rationales for the focus on the multivariate, meta-regression approaches in research question three. In research question three, I shift to an analytic approaches that incorporate many characteristics simultaneously in a model, which makes it much more reflective of the actual experience of study abroad and incorporates all the guiding theory together.

Research Question 3

The third research question is analyzed in four sections. One for each of the three meta-regressions based on the set of moderator variables examined: student, study abroad program, and research design characteristics. These were followed by a forth meta-regression model that combined variables from all three sets. Table 38 provides a summary of the findings from research questions two and three as they relate to the initial hypothesis. In reading the table, + denotes an expected or actual positive association, - denotes an expected or actual negative relationship, ϴ denotes no expected or no actual relationship, a blank denotes the variable was not included in the analysis.

Table 38

Summary of Hypotheses and Findings

Variables	RQ 2		RQ 3		RQ3 Combined	
	Hypothesis	Actual	Hypothesis	Actual	Hypothesis	Actual
Student Characteristics						
Gender	0	-	0	-	0	0
Program of study	0	+	0	+	0	+
Year of study	0	0	0	0		
Ethnocultural Identity	0	0				
International/domestic	0	0				
Previous experience abroad	0	0				
Program Characteristics						
Duration	0	0				
Language of instruction	+	+	+	+	+	+
Faculty instruction	+	+	+	+	+	+
Housing	+	0				
Nature of study abroad group	+					
Cultural interaction	+	0	+	+	+	+
Guided reflection	+	0	+	-	+	-
Research Design Characteristics						
Publication type	+	+	+	+	+	+
Researcher involvement	+	0				
Survey instrument	-	-	-	-	-	-
Research design	+	+	+	+	+	+

The first meta-regression model, focussing on proportion of female students, program of study, and year of study was not significant overall, but the effect sizes have practical implications. A study of exclusively female students was associated with a lower effect size by 0.36. Similarly, a study of only business or only STEM students were associated with an effect about 0.15 or higher than students in social science programs. There was no significant difference for year of study except that students in ‘other years’ score significantly higher than other students.

The second meta-regression focussed on study abroad program characteristics. Recall that no individual moderator variables within this set were associated with mean effect size but using multiple variables in the same, two of them, language of instruction and faculty instruction showed large differences between categories of effect sizes. In the program characteristics meta-regressions, studies in a student' primary languages were associated with a lower effect sizes (g = 0.17) than a study primarily in a foreign language. Moreover, studies that had intentional cultural interaction would be predicted to have effect sizes 0.18 higher, while studies with guided reflection would be expected to have effect sizes 0.17 lower than those without.

Duration was not significant in terms of understanding the mean effect size, and it was not included in the final meta-regression models. This finding suggests that longer is not necessarily 'better' with respect to intercultural competence. Varela (2017) found larger effect sizes for short term programs (g = 0.59) and semester and longer programs (g = 0.51), than for medium-length programs (g = 0.24). Vande Berg et al. (2009) found almost the opposite as the largest effects were for the students who were abroad from 13-18 weeks, and smallest for those who studied longer periods of semester to a year long. I would argue that theorizing, a priori, the non-linear effects found in either the Varela or the Vande Berg findings, would be challenging. Moreover, it is not clear how duration is an indicator of quality over and above the program characteristics.

The findings relating to the language used in coursework provided interesting results as studies partially or primarily in a foreign language were associated with larger effect sizes than programs only in a student's primary language. The finding of a meaningful effect size for cultural interaction (positive) and guided reflection (negative)

contradicts Zarges (2016) and Lemmons (2016) who found no meaningful relationships for either variable. Findings in this study conflict with a major finding of Vande Berg et al. (2009) who found those with a mentor they met very often having a larger effect size than those that said that they never had a mentor (d = 0.06). But Vande Berg did not find a significant difference for intentional cultural interaction (like housing). This study was not able to code with a level of nuance that might allow a more direct comparison to these specific findings from Vande Berg.

Another specific area of note relates to housing as this study found no differences in mean study effect size by housing type. This finding challenges those from Hyndman (2009) who did find that living with host families was significantly related to effect size, and Nichols (2011) who also found that it was positively associated. The Varela (2017) meta-analysis, like the current study, found no significant difference in housing options (with homestay and in dorms with study abroad students).

The third meta-regression focussed on research design characteristics. There are three noteworthy findings. The first is that articles had effect sizes approximately g = 0.14 larger than dissertations. This provides some support for the hypothesis around publication bias as dissertations with "non-significant" or less positive results can be completed, while articles with similar findings may not be submitted, or accepted, for publication in journals.

The second finding was that there was a significant difference in mean effect size by the specific survey instrument used. Studies that used the CCAI and the GMS had effect sizes larger than those that used the IDI or GPI instruments. The final finding in this section was that studies that used a retrospective pre-test had significantly higher

effect sizes than the other two formats. Both findings will be discussed in greater detail below.

The fourth and final model was a meta-regression that used characteristics from each three sets of moderators. The findings from this last analyses found a coefficient for proportion of female students that was $g = 0.09$. This is much lower than in the student characteristics section ($g = -0.36$), and the coefficient has changed from negative to positive. The negative association for proportion of social science students (compared to proportion of Business or STEM students) was consistent. The importance of using a multivariate approach might be most evident in reviewing these variables. In the initial univariate model there was a negative effect size ($g = -0.37$) and a small *p*-value ($p = 0.18$) for the proportion of female students in a study, but in the student characteristics meta-regression model the effect size was similar and negative ($g = -0.36$), and the *p*-value was much larger ($p = 0.62$). In the final integrated model (Table 36), the coefficient for proportion of female students is actually positive ($g = 0.09$), though the *p*-value is quite larger ($p = 0.76$). The rationale for this could be associated with the significant correlation between the proportion of female students in a study and the proportion of social science students in a study. Throughout all the models in the study, the effect size for students in social science was smaller than those in Business or STEM programs. Thus the studies that have larger or primarily social science students also have qualities that are associated with lower effect sizes (such as guided reflection). So the coefficient for female was possibly confounded with either, or both, social science and guided reflection, but once all three variables were in a single model the negative relationship between the proportion of female students and mean effect size disappeared.

In Vande Berg et al. (2009) the authors reported that students identifying as female had statistically significant changes in their intercultural competence (using the IDI), while students identifying as male did not. Though the study did not share effect sizes, using data from the study we would estimate that female students scores improved by approximately $g = 0.25$ and males by approximately $g = 0.05$. Neither Vande Berg et al., nor any other published study abroad research present any rationale why one gender might score higher or lower than another. The finding of lower effect size for social science students challenges a finding in Vande Berg et al., which found significant gains for humanities and social science students, but not other programs (2009). Finally in the Varela meta-analysis, Business students were found to have larger effect sizes than students in other programs of study, a similar finding to this dissertation.

The positive association for course work (whether mostly or primarily) outside of a students' primary language was consistent in this final model as well. Effect sizes for primarily foreign language programs were larger than those primarily in students' first language. This supports findings in Vande Berg that students in programs in a foreign language had larger intercultural gains than those who studied in their own language ($d = 0.25$ versus $d = 0.06$). But, they challenge results in Spenader & Retka (2015) who in their eight sub-samples found some samples where students studying all in English that did have significant changes in their intercultural competence, but also some that did not.

The language finding does lend some support to the Engle and Engle framework, however only 12 studies were of students primarily in a foreign language and about a third (n = 29) had some instruction in a foreign language. The overwhelming majority of the studies in this dissertation are of U.S. students studying in Europe. Thus even when

some coursework was not in their primary language (usually English) it was still quite likely in a Colonial language and one that is not a huge distance culturally or far from their comfort zone. And, as noted in some of the broader critiques of study abroad and intercultural competence in chapter two, that so many students can study abroad in another country having no, or very limited host country language skills reinforces notions of Western privilege. The importance of a local language, especially as it relates to culture, is best understood by noting that the Chinese word for culture is '*wenhua*' (Deng & Tang, 2015). *Wen* can be understood as "to feel and experience the nature by heart" while *hua* means "to change" (Deng & Feng, 2015 p. 145). Combined these characters speak to change through a process of deep empathy, learning and understanding of which language surely plays a significant role. The degree to which this process is possible studying only introductory Germanic or Latin-based languages, which represent the identities of most study abroad participants, is unclear.

Thirdly the findings that studies using the IDI (and the GPI) had lower effect sizes, confirmed the hypothesis. This supports the broad hypothesis that study effect size would be related to survey instrument. Four studies (Anderson & Lawton, 2011; Hansen, 2010, Stromberger, 2010; Williams, 2002) used two instruments on the same population. Using data from Anderson and Lawton, the effect size for the GPI instrument was g = 0.75, while the effect size for the IDI was g = 0.55. These are much larger magnitudes than the median study for those instruments in this dissertation. Williams found larger effect sizes from the Intercultural Sensitivity Instrument (g = 0.92) than the results with the CCAI as (g = 0.15). Varela (2017) also found that studies using the IDI had smaller effect sizes than the median study which aligns with the findings in this study.

The final finding in this section was that studies that used a retrospective pre-test had significantly higher effect sizes than the other two formats. This was consistent throughout research question two and the previous meta-regression models in research question three. Though only two studies used this approach, I argue that if future researchers also use this approach, to exercise considerable caution when interpreting and integrating their findings into the larger research on study abroad as this design produce larger estimates (Groves et al., 2009).

Findings and guiding study frameworks and theory

A guiding theory for this study was Allport's Intergroup Contact Theory (1954). ICT is based on the notion that bringing together individuals of differing backgrounds, beliefs and ideas could reduce prejudice. The study abroad research community has used this theory or framework quite frequently in research (see Medina-Lopez-Portillo, 2004; Lemmons, 2016; Nam, 2011; Savicki & Cooley, 2011) with the idea that a similar approach might act to enhance students' intercultural abilities. The overall finding of this study that study abroad programs are associated with a mean gain of $g = 0.38$ in intercultural competence would generally support the theory. Many of the variables in the study, particularly the study abroad program variables (including housing, duration) were not found to be significantly associated with larger effect sizes. Only language of instruction, and cultural interaction produced meaningful (positive) effect sizes to support the theory. If it is the 'contact' that reduces prejudice or enhances intercultural competence in this study, how can we understand that levels of the other Engle and Engle variables that would indicate greater, closer, or longer contact were not associated with larger study effect sizes?

Allport's theory (1954) is based on four principles: common goals, intergroup cooperation, no competition, and authority support. The theory posits that controlling and optimizing for these four factors would be associated with the desired changes. What would be optimal? According to the Engle and Engle theoretical framework, it would be a program where there was a lot of intentional interaction, meaningful guided reflection, housing with local students or families, longer durations, and teaching in the local language. The large number of 'optimal' and related factors may provide a reason why so few statistically significant, individual relationships were found in the analysis for the study abroad program characteristics. While the dissertation used many individual studies, overall this is not a highly powered study (n = 85) and it is possible that some multicollinearity is present. Multicollinearity occurs when independent variables are correlated (Borenstein et al., 2009). However, correlation tables did not indicate that there were many situations where independent variables were correlated greater than $r = 0.30$.

Pettigrew and Tropp (2006) and more recently Lemmons (2016), focus on the difference between intergroup contact and intergroup proximity. While students abroad are in proximity to the conditions for intercultural or intergroup contact, the degree to which contact occurs is less well known. Few studies provided detailed information about the degree of intercultural interaction or the types of activities in which students engaged with local communities. This is partially why it was decided to reduce both the intentional cultural interaction and the guided reflection variables to a very simple yes/no variables as few studies provided necessary details relating to these moderators. Rarely did subsequent contact with authors provide enough details to meaningfully code data on

a more nuanced level, only enough information to confirm whether any of these activities were features of their programs.

Referring back to the Engle and Engle framework, there is similarity with the Intergroup Contact Theory, in looking at the notion of optimal level of contact. The highest levels of their framework would have a program at least a semester long, in the host language, directly enrolled in local courses, regular participation in cultural integration and extensive reflective exercises. The number of studies that met these criteria is likely quite small. It is possible that some of these factors in the Engle and Engle model interact with each other. But my ability to test this hypothesis would require a considerable larger sample of highly detailed studies and/or a focus on a smaller number of categories in the classification system.

It could also be argued that the framework is lacking in at least three dimensions. The first is that the Engle and Engle framework does not include space for pre-departure preparation and training. These are initiatives to help prepare students for their intercultural experience before they begin, which while it might impact their pre-test measurement, might also help more students be better prepared to engage, immerse and learn while abroad. A second variable is on-site or in-country orientation. This variable could plausibly be included in the guided reflection, though I would argue it might merit its own category. A successful orientation would introduce a student to their new location, program and procedures. It could highlight the resources and supports available to students, beyond intercultural development, but to aid with challenges and concerns similar to those they may have faced at home. A third omission is ongoing support. This is not directly focused on intercultural competence, but with daily or ongoing challenges

of housing, academic accommodations, health and wellness. These supports could be provided by staff, student clubs and organizations, or faculty. Students are still likely experiencing these concerns while abroad, just in a new cultural context.

Overall the ability to consider the Engle and Engle framework and through it Allport's Intergroup Contact Theory, using a more nuanced approach considering both how often and how frequently these intercultural interactions occur, might provide a more robust and valid approach to testing the theories. Both provide an excellent starting point for analysis and coding, but the results of this dissertation provide only weak to moderate support that they further our understanding of what does and does not contribute to greater intercultural competence development.

A final guiding framework in this study relates to validity. Of the four primary threats to internal validity, two, maturation and selection, are the biggest concerns in the research reviewed for the study. Maturation is the notion that individuals change over time. This is what necessitates, at minimum, a comparison group. One of the criteria for this study was that all studies must use at least two measures of intercultural competence. This contrasts with the Varela (2017) study that used numerous between-group comparisons; effectively a comparison of a post-test of a study abroad group and a post-test of comparison group. Without knowing the initial measure of intercultural competence, it is not possible to know a) if there was change and b) if that change was greater for one group than another. Pre-test/post-test designs help in this regard as it allows researchers to compare changes within a group over time. Moreover, based on my reading of the studies in the dissertation, none were based on students in community college, all were university students. And, as noted earlier the participants are

overwhelmingly from U.S institutions studying abroad in Europe or other Western countries. Thus the degree to which these findings are relevant for students, staff and institutional leaders in community college contexts or for those outside of Western countries is limited.

Limitations to the study

A meta-analysis is an attempt to synthesize a large body of research into some common findings that help advance knowledge from the cumulative and combined power of many researchers' efforts. However, like all forms of research, this study has numerous limitations. Five of them are detailed below.

Undercoverage. The first limitation is the threat of undercoverage. It is improbable that this study represents the universe of studies that meet the four study criteria. These were that a study was published in English, that it focuses on a for-credit study abroad program, that the outcome is associated with intercultural competence and that the measurements are quantitative. While extensive searches have been made of databases, hand searches of journals, forward and backwards searches of reference lists, undoubtedly studies have been missed. The possibility of undercoverage became apparent when seeing how many studies were found via forwards and backwards searches. The two approaches to understanding impacts of undercoverage, classic failsafe (Rosenthal, 1979) and Orwin's fail-safe N (1983) both suggest that it would take a large number of studies to significantly change the overall study finding. And the impact of undercoverage is somewhat minimized by having a large, at least for a meta-analysis, number of coded effect sizes. The bigger concern would be if there was systematic undercoverage that could impact the relationships (both significant and not) in research

question three. It should also be noted again that this study did not search for any studies published after December 21, 2016. It is possible that a few dozen additional studies could have been identified and coded which would have provided greater statistical power for identifying moderating effects. Undercoverage also affects the power of the study. A larger number of studies would permit a more robust analysis of moderators and potential interactions in meta-regression analysis. The use of the Engle and Engle framework also potentially contributed to limiting the number of studies in the dissertation that were based outside of the U.S. or included non U.S students. I chose a series of U.S.-centric frameworks, which, resulted in an overwhelming U.S.-centric sets of studies for analysis. A future research effort focused on studies of primarily students from Europe, within Europe, or other geographies where the research literature can be combined would be an excellent companion to this study.

Secondary coders. Secondary coders reviewed twenty-six of the effect sizes used in the study. A meta-analysis or systematic review published through either the Campbell Collaboration (Social Sciences) or Cochrane (Medicine), all included studies have a secondary coder. This is the ideal scenario. However, given the costs involved, only about 30% of the studies in this dissertation had a secondary coder. Still, having secondary reviewers in a social science meta-analysis dissertation may be more the exception than the rule. In a review of fifteen meta-analysis on college students or issues relating to post-secondary students education, only three noted that any of their studies had a secondary coder and the remaining did not mention it was a possible limitation. Nevertheless, additional secondary coders likely would improve the quality of the coding in the study, minimize error and enhance validity.

Duration between pre-tests and post-tests. A third limitation relates to the coding of the study abroad program characteristics. While duration of the program was coded and analyzed, the pre and post survey times do not always correspond to those same time frames. I would have preferred to include the time between the pre and post-test not the actual duration of the study abroad program. Several studies noted that the pre or post-test survey was open to students for several weeks or months, meaning that a program with a duration of 4 weeks, might have a real pre-post survey duration of 8 weeks or more. The inability to code this accurately is a limit as there could be an intersection between duration of a program and the actual timing of the surveys. This approach would create a rule for eliminating the studies that used retrospective pre-tests, as the time between the pre- and post-tests would be zero regardless of how long the program of study is.

Quality of study abroad program variables. The Engle and Engle framework provided an idealized, but ultimately over-ambitious guide for coding study abroad characteristics. Some variables like duration, language use, faculty instruction and even housing, were coded at multiple levels, but as noted before, two of the most important variables, intercultural contact and guided reflection for analysis were not. This was due to how few studies provided any details on these variables and how few studies provided enough details to meaningfully code them in a nuanced way. Without enough details, or if a study provided no details at all, I chose a conservative approach, and artificially dichotomized them. A potential impact of this is that it might minimize the mean effect size if both variables were positively related to intercultural competence or if studies that had higher levels of either intercultural contact and, or, guided reflection, variables were correlated with mean effect sizes.

Omitted variable bias. Omitted variable bias refers to items that could be correlated with the outcome of interest and one or more independent variables (Vogt, 2007). Three relate to student identity variables: first-generation status (whether a student had a family member with previous post-secondary experience), socio-economic status of the student, and whether the students were using any type of financial aid. However, none were consistently present in the coded studies. These variables, in addition to ethno-cultural/racial identity, address some of the concerns scholars have of inequities of the profile of study abroad participants and who is more inclined to participate and benefit (see Stroud, 2010; Twombly et al., 2012).

Another variable in study abroad program variables was what Hofstede (1983) called cultural distance; a measure of the 'distance' between the communication styles, preferences and approaches of two countries. This is used as a proxy in study abroad research for how different, or foreign, a country is. The variable was not used for two reasons. First, some studies (see Engberg & Jourian, 2015; Nichols, 2011) included students studying in many different countries or from many different countries. Without student level data it would not be possible to assign a cultural distance value to these studies. Secondly, there are a growing number of international students studying abroad. The use of a home country value for someone who might have only been in the country for a year or two (in the case of an international student or recent immigrant) and is potentially studying abroad in a country that is more like their country or origin, is an analytic challenge I can not solve or devise a defensible approach to.

A final set of omitted variable relate to study abroad motivations or perhaps more interestingly, what Engberg and Jourian (2015) call intercultural wonderment. In both

cases these are measures to understand either why a student wants to study abroad and what their goals and intentions are (Burrow, 2010; Kitsantas, 2004) or the students willingness to engage, explore and push themselves outside of their typical comfort zone (Engberg, Jourian, & Davidson, 2016). Engberg et al. have found that intercultural wonderment mediated the development of intercultural competence. While both motivation and wonderment are potentially interesting and important constructs, among the studies included in the analysis, only a handful used any sort of measurement that speaks to these constructs.

Study Implications and Recommendations

This section focuses on implications and recommendations for study abroad program staff (those choosing, planning and leading study abroad programs) and/or faculty who lead study abroad programs, for institutional leaders and policy makers that determine policy relating to study abroad and finally implications for researchers of study abroad and meta-analytic researchers.

Implications and recommendations for study abroad program staff. For those who select or recruit study abroad participants the major implication of the study is that there is no 'ideal' group of students to study abroad. While a small negative effect was found for the proportion of female and the proportion of social science students in a study, they are meaningful at a study level only if the study was approaching exclusively all female or all social science students. The study did not find an effect for composition by year of study or previous experience abroad or international/domestic status, so if a program director included some second year students in a program of largely fourth year students, there is no evidence that this impacts the overall group experience in terms of

intercultural competence development. There may be good reasons why a student, possibly in a very strictly designed academic program of study or a student on academic probation, or someone without previous experience abroad might not want to study abroad, or may be cautioned in their application, but it should not be because they are less likely to benefit from the experience than others. If there are criteria that prohibit certain populations from applying, these procedures should be discussed and re-examined.

Implications and recommendations for faculty and study abroad program designers. The major finding for this group was that there were few significant program variables related to the effect sizes in the study. This does not mean that program design does not matter, but notions that a learning environment for intercultural competence is best designed by the upper levels of the Engle and Engle framework are only moderately supported. There is some support for programs in a foreign language, faculty instruction, and some support for including intentional cultural interaction (though unexplained small negative relations with guided reflection) but the evidence for true best practices in program design remains elusive.

Secondly, while additional analysis that grouped weeks of study into short, medium and long-term could provide different results like those found in Varela (2017) and Vande Berg (2009), the absence of a relationship between intercultural competence and duration, should be an opportunity for staff and faculty to consider developing more short-term opportunities. Though longer programs may be idealized by staff and administrators, and more reflective of ones they participated in as students, they may also come with larger costs (housing, lost opportunities for income at home) and greater

potential challenges for course credit transfer which may create additional barriers for students to participate.

A third implication relates to faculty who lead study abroad programs also designing research studies. Many studies used an analytic technique that does not support their conclusions. The most common example was conducting *t*-tests on a group abroad and a group at home, then concluding that the study abroad group changed more, if it had a significant pre-post difference and the at home group did not. In designs like this it is more appropriate to use ANCOVA or regression approach which can account for the changes within and between the groups over time. Many of these issues can be addressed with teams conducting studies using more rigorous research designs. It is also incumbent on journal editors to identify reviewers who have the background necessary to highlight these weaknesses in manuscripts before publication.

Faculty who lead study abroad programs and wish to do quantitative research studies should ensure they have a research team experienced in quantitative analysis. There are numerous examples in conclusions of authors saying the significant result in their study 'proves' that study abroad programs enhance intercultural competence. The use of causal language may not be intentional, but it is widely present. Moreover, the number of times that *p*-values were compared as evidence of one finding being 'more' or 'less' significant than another were widely present. While there are no perfect methodological approaches, an error seen repeatedly is a researcher having a group of students abroad and a group at home and comparing whether the changes for each group, independently were significant or not, rather than integrating all participants into a study.

Implications and recommendations for institutional leaders, policy makers and lobby groups. Considerable attention in the first two chapters of this dissertation focussed on how levels of government, lobby groups and institutional leaders are arguing for dramatic increases in study abroad participation rates. The degree to which their rationale centers on enhancing intercultural competence varies, but it may be worthwhile to focus more on the ways that intercultural competence might be enhanced, outside of study abroad. Bowman (2010) in a meta-analysis on the relationship of cognitive development and diversity experiences at home found a small effect ($g = 0.25$). The effect is smaller than found in this study, but still positive.

If intercultural competence remains a major goal of leaders and policy makers, then consideration of a broader range of opportunities, domestic and international is warranted. Other opportunities like international research, co-operative education/internship or service-learning opportunities might also be explored. Niehaus and Crain (2013) found numerous positive outcomes for students participating in domestic and international service-learning programs and as faculty-led programs these help mitigate concerns about course transfer or time to graduation that are often present in study abroad programs (Hamir, 2011). Engberg (2013) also found positive development for both domestic service-learning programs and international study. In addition with cost being a primary concern in limiting who considers studying abroad, international internships or co-op opportunities may be a way for students to participate in an intercultural learning experience while also earning money.

Secondly the finding that student characteristics were not strongly associated with changes in intercultural competence, should lessen concerns leaders have about investing

funds to support more students from under-served backgrounds to participate. In the Open Doors data more than 70% of study abroad participants identify as White, while under 60% of the U.S. population of post-secondary students do (IIE, 2017). Policies which might increase participation from under-served populations of students, should be strongly considered. The question of how to increase participation from under-represented groups is unfortunately even less clear than the question of what contributes to intercultural competence.

Thirdly, in thinking about the degree of interest and potential participation in study abroad, the National Survey of Student Engagement (NSSE) provides a useful standardized approach with their focus on high-impact practices (HIPS). HIPS include learning communities (block classes), service-learning, undergraduate research projects, internships or co-ops, study abroad, and capstone experiences. They are considered high-impact because they involve considerable time commitment, interaction with faculty and peers, and use skills from a wide range of academic and co-curricular experiences (Kuh, O'Donnell & Schneider, 2018). Reading the U.S. NSSE data, even at the smallest of institutions (who tend to have the highest study abroad participation rates) the highest study abroad participation rates are around 20%, compared to over 50% of students who participate in service-learning and internships (NSSE, 2018). At the University of Toronto, more than half of graduating students report doing research with faculty, compared to approximately 12% who had studied abroad (University of Toronto, 2017). Despite broad public, lobby group and business community, and institutional leader support, study abroad remains the least participated in of these practices. Greater focus on addressing what motivation and benefits students see in study abroad and what barriers

and misconceptions students associate with study abroad, especially among those who do not traditionally participate, is required for meaningful increases in overall study abroad participation to occur.

Next, I would argue that there is considerable opportunity for groups like CBIE in Canada, NAFSA or FEA in the U.S., to provide leadership, coordination and some funding support to develop needed large scale multi-institutional studies. The oft-cited Georgetown Consortium Study (Vande Berg et al., 2009) is approaching 10 years old and remains one of the best individual papers on study abroad. Still, this is a decade ago. It was published at a time when few students had smart phones with easy and affordable access to friends and family back home while they are abroad, and is arguable that a new, or ideally several large-scale research efforts could add considerable value. This study reviewed some very thoughtful dissertations (most notably Lemmons, 2016 and Nam, 2011), but the GSC remains a study cited in nearly every study abroad paper. These lobby groups and policy makers could support research by encouraging a deeper look at the Engle and Engle program framework, by incentivizing institutions to include these variables in their data collection plans, supporting overall and institutional level analysis and ultimately dissemination.

The 2019 funding announcement by the Canadian federal government of $150,000,000 for international education, presents an unprecedented opportunity for research and evaluation. A joint effort by three sets of stakeholders CBIE, Universities Canada or Colleges and Institutes Canada might build off three elements in this study. The first is to ensure that this opportunity is used to establish a process to systematically collect data about who is participating in all forms of international education, specifically

the work and study abroad opportunities funded through this initiative. This represents the I in Astin's IEO model and at a minimum a focus on collecting data found in Table 9 (gender, program of study, year of study, ethnocultural/racial identity) that Open Doors collects annually. However, I would argue that they should extend that data collection towards broader notions of who is and who is not participating in international education by the inclusion of questions around first in the family to attend post-secondary education, financial aid and socio-economic status.

Secondly, the group could work with institutions to help develop a standard research and evaluation protocol. As project sponsors they could nudge institutions towards using specific instruments (likely the IDI and GPI) and encourage them not to develop their own. These groups could also advocate that work or study abroad programs include some additional language instruction and possibly an intentional focus on pre-departure training, onsite orientation and ongoing support. They could also consider adding the intercultural wonderment scales developed by Engberg and Jourian and encourage some exploration of how intentional cultural interaction and guided reflection might be enacted while abroad.

Finally they could play a key role in the evaluation of the government's funding initiative either directly or by contracting the data analysis to faculty researchers. This would ensure consistent and appropriate analytical approaches were used in the data analysis but would also add validity in having a third party, rather than the participating institutions, analyze the data. They could also use an approach like the National Survey of Student Engagement wherein each institution receives a copy of its own data, plus overall summaries of findings. But the coordinating group also can use the full data set

benchmarking (participants, program design features and outcomes) to do broader analysis using the IEO multivariate approach used in this study. This would represent a dramatic increase in the quantity and quality of study abroad and international research in Canada.

Implications and recommendations for study abroad researchers. There are three major findings for study abroad researchers. The first is associated with developing a deeper understanding of the components, the syllabus really, of the study abroad program in the research. The lack of details about the study abroad program shared in most articles, especially in contrast to the amount of space often devoted in an article or dissertation to the rationale for study abroad or a description of participation is puzzling. In dissertations, including course or program syllabi should be possible. While in journal articles, links are often made online to additional tables and figures. Likely, some edits might be required to ensure the privacy of participating research institutions, but a thorough discussion of what the study abroad program was would be a significant step forward in understanding study abroad program design. This recommendation is most likely to be realized in journals and publications that have an explicit focus on study abroad (such as *Frontiers: Journal of Study Abroad*) or international education.

A second implication questions researchers who act as both a leader of a study abroad program and a lead researcher on a resulting publication. The differences were not significant, but studies in which the researcher was a participant in the study abroad program had slightly larger effect sizes than those that did not. This is not an accusation of misconduct, but the small relationship persisted through different analytic models. A

recommendation might be that faculty who act in this dual role should be explicit in their publication with a potential disclosure statement.

A third implication and recommendation involves comparison groups. Only about a quarter of the studies in this meta-analysis had them and there was almost no difference in the mean study effect size between pre-test/post-test designs and pre-test/post-test with comparison group designs. This finding is puzzling. There is a genuine question about what makes a good comparison group. Some studies used classes of students in 'cross-cultural' courses (psychology, business and social work seemed to be common examples) at home, to help approximate the experience students were having abroad. Given the relatively small proportion of students who do study abroad, this would seem to be a more accurate comparison group, than a group of students who had no interest in study abroad. An important question is whether study abroad is a better intercultural learning opportunity than the other diversity experiences a student may have at their own institution or in their own country (see Bowman 2010; 2011, Denson & Bowman, 2013). Regardless of who are chosen for the comparison group, a design with a comparison group is more rigorous and has greater internal validity than one that does not.

Overall this study found that participation in study abroad had a small and statistically significant relationship with increased intercultural competence. What this study did not find was a large number of student or study abroad program characteristics that were related to these changes. Some of these results were hypothesized, others were not. Summarizing some of the previous sections about implications and recommendations I see three primary reasons for why this study may not have found clearer results.

The first relates back to the limitation of omitted variables, both from a student identity framework (equity variables like first in the family to attend post-secondary education, socio-economic status and financial aid), but also the lack of measurement or variables relating to motivation, goals for study abroad or the interesting concept of intercultural wonderment introduced by Engberg and Jourian (2015). Secondly, the study found significant differences in effect sizes by instrument type suggesting that the conceptualization and operationalization of surveys purporting to measure intercultural competence requires considerably more investigation from a validity framework. The decision to artificially dichotomize both the intercultural interaction and the guided reflection variables reduced both the variance and the nuance in these variables. It is arguable that students can engage in a wide range of meaningful intercultural contact with or without the support of staff and faculty. However this is less likely true for guided reflection. So intentional, authentic, relevant approaches to guided reflection were combined with poorly conceived, poorly supported and executed approaches with could supress the relationship with intercultural competence. Guided reflection requires intentional planning and effort from faculty and staff, and it is possible different results, for both of these variables, could have emerged if those nuances were coded.

Finally, as discussed when reviewing implications relating to the Engle and Engle framework and Allport's Intergroup Contact theory, there remains much to learn about the process of developing interculturally. We can approximate contact, or at least proximity, but the process of intercultural competence development is less clear. Furthermore, it should be remembered that all the instruments and frameworks for the dissertation emanated from Western contexts. Allport's theory speaks to conditions that

support or potentially make intercultural development more likely but focuses much less on the process by which such development occurs. To better understand this developmental process, scholars and practitioners might return to the models of intercultural development reviewed in chapter three and also take a more comprehensive look at the findings from mixed methods and qualitative studies on study abroad and intercultural competence. They might also consult literature of study abroad participants outside Western countries.

***Varela meta-analysis*.** Though the overall effect size is very similar to that in the Varela meta-analysis, this study has three key differences or advantages over it. First, the current study only includes studies that used, at minimum, a pre-test/post-test design, while the Varela study used numerous studies that were cross sectional, which has significant internal validity issues. Among the 30 studies Varela included were several that used a between-groups studies design. In those studies a measurement was taken at the end of a period for both a study abroad and a group at home and the results are compared. However, this approach fails at least one major threat to internal validity. The first is selection as individuals develop and change over time. Without a pre-test of the abroad group it is impossible to rule out the possibility that either of the groups were always different from each other or that they changed at all while abroad. Between-groups studies can be used if participants are both randomly selected into the study and randomly assigned to the experiment (study abroad) or comparison (at home) groups, but this design approach seems challenging for experiential learning programs like study abroad.

Secondly, this dissertation included 89 different effect sizes compared to just 38 in Varela. It is not clear why there is such a difference, though it appears Varela has significant undercoverage for dissertations and there may be a systematic difference between our studies. While it could be argued that dissertations as a product, and graduate students as researchers, are not the same quality as journal articles published by faculty, this ignores the fact that the dissertations represent studies of real students participating in real study abroad programs. Leaving out dissertations enhances possible publication bias. Moreover, the magnitude of the difference in the number of studies used between the two meta-analyses is even greater when considering that several studies in Varela were between-groups designs.

Finally, the study was conducted using a more rigorous analytical and theoretically informed approach than Valera who only used the individual moderator approach found in research question two. Without a multivariate approach, potential suppressor effect are not examined. The multivariate approach also recognizes the complexity of the study abroad experience and the intercorrelated relationships of the variables in the model. Organizing the study, and analysis, around Astin's IEO theory broadly and Engle and Engle's specifically provides a more defensible approach to the analysis and enhances the validity and generalizability of the findings.

Conclusion

I came to this specific dissertation topic after planning on doing original data collection on study abroad. Over time, I became frustrated reading articles about study abroad that concluded in one of two ways. The first was a specific claim that future research should investigate various items that could potentially moderate or enhance the

intercultural competence development of students abroad. The second was frequent references to the Engle and Engle framework, suggesting it could provide a useful framework for conducting a study and analyzing data. So, I changed my focus from primary to secondary data collection and analysis. This dissertation is the largest attempt to synthesize and quantify a relationship between study abroad and intercultural competence.

The literature on study abroad has grown considerably since this dissertation began and more clarity on the research questions raised in this study would have been found by including studies beyond the cut-off date (December, 2016). I look forward to future researchers updating, expanding, critiquing and improving this study using more studies, more detailed criteria and more advanced analytic methods and I am happy to share this data set to support them in that work. My perspective on study abroad as an academic endeavour has evolved considerably since beginning my doctoral program. When I began, study abroad was the area I had, and thought I would continue to, work in. Over time, my interest in meta-analysis grew from the logical approach to the study I wanted to conduct, to an area of genuine interest, fascination, and at times incredible frustration, especially by the technical complexities and possibilities in it.

The growth of study abroad research, especially in dissertations, is an exciting and positive development for all sets of stakeholders. This work, however, is only useful to the extent that study abroad program officers, faculty who lead programs, institutional leaders and policy makers and researchers can review, evaluate, synthesize and consume it. This challenge is not unique to those in the field of study abroad; but the lack of faculty whose research programs are focused on these topics exacerbates the challenge.

Few researchers have published more than one article on study abroad. This dissertation is a small attempt to aid these stakeholders in understanding the complex, and still largely not well understood, relationship between study abroad and intercultural competence.

Appendix A: Bibliographic Details

Descriptor	Instructions and Examples
Date Found	Date the study was identified as being included/excluded from study
Date Analyzed	Date study was coded
2nd coder	Date study was coded by second coder if available
Method found	Method in which the study was identified
Title	Name of the study
First author	Enter the name of the primary author (last name, initials). Example (Burrow, J.)
Second author	Enter the name of the second author (last name, initials).
Third author	Enter the name of the third author (last name, initials).
Author contact	email addresses of authors
Year	Enter the publication year of the study.
Publication Type	Enter the publication type of the study 1= Journal Article/conference paper 2=Dissertation or thesis
Citation	Enter full APA citation of the study
Faculty/Research	Are the authors also involved in the study abroad program?

Appendix B: Student Characteristics

Descriptor	Instructions and Examples
GTx	Gender proportion of abroad group expressed as % female
BusTx	Proportion of abroad students in Business Programs
SocTx	Proportion of abroad social science students
SciTx	Proportion of abroad STEM students
Otx	Proportion of abroad other students
Y1Tx	Proportion of abroad students in first year
Y2Tx	Proportion of abroad students in second year
Y3Tx	Proportion of abroad students in third year
Y4Tx	Proportion of abroad students in fourth year
YOTx	Proportion of abroad students in other year
YG	Proportion of abroad students in grad program
DomTx	Proportion of abroad students as domestic
IntlTx	Proportion of abroad students as international
BNHTx	Proportion of abroad students who are African American
AIANTx	Proportion of abroad students who are American Indian/Alaska Native
APITx	Proportion of abroad students who are Asian/Pacific Islander
HTx	Proportion of abroad students who are Hispanic/Latinx
WNHTx	Proportion of abroad students who are White
REUTx	Proportion of abroad students whose race/ethnicity is not known or identified
PATx	Proportion of abroad students who have been abroad before
PNATx	Proportion of abroad students who have not been abroad before
GC	Gender proportion of comparison group expressed as % female
BusC	Proportion of comparison group students in Business Programs
SocC	Proportion of comparison group social science students
SciC	Proportion of comparison group STEM students
OC	Proportion of comparison group other students
Y1C	Proportion of comparison group students in first year
Y2C	Proportion of comparison group students in second year
Y3C	Proportion of comparison group students in third year
Y4C	Proportion of comparison group students in fourth year
YOC	Proportion of comparison group students in other year
YG	Proportion of comparison group students in grad program
DomC	Proportion of comparison group students as domestic students
IntlC	Proportion of comparison group students as international student
NRAC	Proportion of comparison group students who are non resident aliens
BNHC	Proportion of comparison group students who are African American
AIANC	Proportion of comparison group students who are American Indian/Alaska Native

APIC	Proportion of comparison group students who are Asian/Pacific Islander
HC	Proportion of comparison group students who are Hispanic/Latinx
WNHC	Proportion of comparison group students who are White
REUC	Proportion of comparison group students whose race/ethnicity is not known or identified
PAC	Proportion of comparison group students who have been abroad before
PNAC	Proportion of comparison group students who have not been abroad before

Appendix C: Study Abroad Program Characteristics

Descriptor	Instruction and Examples
Dur	Duration of time abroad expressed in weeks
LOI	Language of instruction. 1 = All in students primary language, 2 = Mostly in students primary language except for introductory language in second language and 3 = Primarily or totally in a foreign language
CAW	Instruction of study abroad program. 1 = Program taught by home professors, 2 = Program taught by combination, 3 = Program taught by local faculty
Hou	Student Housing arrangements. 1 = homestay, 2 = accommodation with fellow study abroad students, 3 = accommodation with other study abroad or international students, 4 = accommodation with local students
GuiSCI	Guided opportunities for cultural interaction. This is a description of the out of class opportunities the program provides and rated 1-5 to strongly integrative
Ref	Opportunities for reflection either through service learning projects, on site or online mentors

Appendix D: Research Design Characteristics

Descriptor	Instructions and Examples
StTy	Study Type. 1 = pre-post study abroad only, 2 = pre-post study abroad and pre-post comparison group, 3 = retrospective pre-test
Instr	Instrumentation. 0 = non-standard instrument, 1 = Standardized instrument
InstrTy	Specific Instrument 1 = GPI, 2 = IDI, 3 = CCAI, 4 = BEVI, 5 = GMS, 6 = MGUDS, 7 = Other
Group/Block	Nature of study abroad group. 1 = collective group(s), 2 = individual abroad student/small groups

References

Abrams, I., & Hatch, W. R. (1960). *Study abroad: New dimensions in higher education*. Washington, D.C.: U.S. Office of Education-Government Printing Office.

Adams, T., Leventhal, M., & Connelly, S. (2012). International student recruitment in Australia and the United States: Approaches and attitudes. In D. Deardorff, H. de Wit, J. D. Heyl, & T. Adams (Eds.) *The SAGE handbook of international higher education.* (pp. 399-417). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397.n22

Advisory Panel on Canada's International Education Strategy [APCIES]. (2012). *International education: A Key driver of Canada's future prosperity*. Ottawa, ON: Foreign Affairs and International Trade Canada.

Allport, G. W. (1954). *The Nature of Prejudice.* Cambridge, MA: Addison-Wesley.

Altbach, P. G. (2012). The globalization of college and university rankings. *Change: The Magazine of Higher Learning*, *44*(1), 26-31. doi:10.1080/00091383.2012.636001

Altbach, P. G., & Knight, J. (2007). The internationalization of higher education: Motivations and realities. *Journal of Studies in International Education, 11*(3-4), 290-305. doi:10.1177/1028315307303542

Altbach, P. G. (2006). Globalization and the university: Realities in an unequal world. In J. J. F. Forest & P. G. Altbach (Eds.), *International handbook of higher education* (Vol. I, pp. 121–140). Dordrecht, The Netherlands: Springer.

Amani, M., & Kim, M. M. (2018). Study abroad participation at community colleges: Students' decision and influential factors. *Community College Journal of Research and Practice*, *42*(10), 678-692. doi:10.1080/10668926.2017.1352544

* Anderson, P. H., & Lawton, L. (2011). Intercultural Development: Study abroad vs. on-campus study. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *21*, 86-108.

* Anderson, P. H., Lawton, L., Rexeisen, R. J., & Hubbard, A. C. (2006). Short-term study abroad and intercultural sensitivity: A pilot study. *International Journal of Intercultural Relations, 30,* 457–469. doi: 10.1016/j.ijintrel.2005.10.004

Angulo, S. K. (2008). *Identity change in students who study abroad* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3311181)

* Armfield, P. A. (2004). *An examination of the relationship between students' interaction with host nationals while on study abroad and their development of*

intercultural sensitivity (Doctoral dissertation). Retrieved from https://drum.lib.umd.edu/handle/1903/1450

Association of American Colleges, & National Leadership Council (US). (2007). *College learning for the new global century: A report from the National Leadership Council for Liberal Education & America's Promise*. Association of American Colleges.

Association of Universities and Colleges of Canada. (2007). *Canadian Universities and International Student Mobility*. Retrieved from http://www.aucc.ca/wpcontent/uploads/2011/05/student-mobility-2007.pdf

Astin, A. W. (1970). The methodology of research on college impact, part one. *Sociology Education, 43*(3), 223-254.

Astin, A. W. & Antonio, A. L. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Lanham, MD: Rowman & Littlefield Publishers.

Australian Government. (2017). *National Strategy for International Education 2025.* Retrieved from: https://nsie.education.gov.au/sites/nsie/files/docs/national_strategy_for_international_education_2025.pdf

Baird, L. L. (1973). Teaching styles: An exploratory study of dimensions and effects. *Journal of Educational Psychology, 64*(1), 15-21. doi:10.1037/h0034058

Banks, M., & Bhandari, R. (2012). Global student mobility. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 379-399). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/978142218397.n21

Barrow, C. W., Didou-Aupetit, S., & Mallea, J. (2003). Globalisation, Trade Liberalisation, and the Knowledge Economy. In *Globalisation, Trade Liberalisation and Higher Education in North America* (pp. 1-21). Dordrecht, Netherlands: Springer.

* Bates, J. T. (1997). *The effects of study abroad on undergraduates in an honors international program*. (Unpublished doctoral dissertation). University of South Carolina, Columbia. doi:10.1086/224504

Becker, T. (1968). Patterns of attitudinal changes among foreign students. *American Journal of Sociology, 73,* 431-442.

Belkhodja, C., & Esses, V. (2013). *Improving the Assessment of International Students' Contribution to Canadian Society.* Knowledge Synthesis. Retrieved from

http://p2pcanada.ca/wp-content/uploads/2014/02/International-Students-Contribution-to-Canadian-Society.pdf

Bennett, J. M. (1986). Modes of cross-cultural training: Conceptualizing cross-cultural training as education. *International Journal of Intercultural Relations, 10*(2), 117-134. doi:10.1016/0147-1767(86)90002-7

Bennett, J. M. (2010). A short conceptual history of intercultural learning in study abroad. In W. Hoffa & S. DePaul (Eds) A history of U.S. study abroad: 1965-present. Special publication of *Frontiers: The Interdisciplinary Journal of Study Abroad,* (pp. 419-449). Carlisle, PA: Forum on Education Abroad.

Bicknese, G. (1974). Study abroad Part I: A comparative test of attitudes and opinions. *Foreign Language Annals, 7*(3), 325-336. doi:10.1111/j.1944-9720.1974.tb02592.x

Black, G.L. and Bernardes, R. (2014) Developing global educators and intercultural competence through an international teaching practicum in Kenya. *Comparative and International Education, 43(*2). Retrieved from https://ir.lib.uwo.ca/cie-eci/vol43/iss2/4

Black, H. T., & Duhon, D. L. (2006). Assessing the impact of business study abroad programs on cultural awareness and personal development. *Journal of Education for Business, 81*(3), 140-144. doi:10.3200/JOEB.81.3.140-144

Black, J. S., & Mendenhall, M. (1991). The U-curve adjustment hypothesis revisited: a review and theoretical framework. *Journal of International Business Studies, 22,* 225-247. doi:10.1057/palgrave.jibs.8490301

Blomeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C., & Fege, J. (2013). *Modeling and measuring competencies in higher education*. Rotterdam, Netherlands: Sense Publishers.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York, NY: McGraw-Hill.

* Boatler, R. W. (1992). Worldminded attitude change in a study abroad program: Contact and content issues. *Journal of Teaching in International Business,* 3(4), 59-68. doi:10.1300/J066v03n04_04

Bond, S., & Lemasson, J. P. (Eds.). (1999). A *new world of knowledge: Canadian universities and globalization.* Ottawa, ON: International Development Research Centre.

Bond. S., Girgrah, A., Burrow, J., Vander Meulen, C., Spaling, M., & Areepattamannil, S. (2009). *World of learning: Canadian post-secondary students and the study abroad experience.* Ottawa, ON. Canadian Bureau of International Education.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley-Blackwell.

Bowden, J., & Marton, F. (1998). *The University of learning: Beyond quality and competence in higher education.* London, England: Kogan Page.

Bowman, N. A. (2010). College diversity experiences and cognitive development: A meta-analysis. *Review of Educational Research, 80*(1), 4-33. doi:10.3102/0034654309352495

Bowman, N. A. (2011). Promoting participation in a diverse democracy a meta-analysis of college diversity experiences and civic engagement. *Review of Educational Research, 81*(1), 29-68. doi:10.3102/0034654310383047

Bowman, N. A. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education, 53*(3), 375-382. doi:10.1007/s11162-011-9232-5

Braskamp, L.A., Braskamp, D. C., & Engberg, M.E. (2013). *Global Perspective Inventory: Its Purpose, Construction, Potential Uses, and Psychometric Characteristics*. Retrieved from https://gpi.central.edu/supportDocs/ manual.pdf

Braskamp, L. A., Braskamp, D. C., & Merrill, K. C. (2008). *Global perspective inventory (GPI): Its purpose, construction, potential uses, and psychometric characteristics.* Chicago, IL: Global Perspective Institute Inc. Retrieved from https://gpi.central.edu/supportDocs/manual.pdf

Braskamp, L. A., Braskamp, D., & Merrill, K. C. (2009). The value added of education abroad: Its impact on global learning and development. *Frontiers: The Interdisciplinary Journal of Study Abroad, 18*, 101-118. Retrieved from https://eric.ed.gov/?id=EJ883693

Brennan, M.G. (Ed.). (2004). *The origins of the grand tour*: *The travels of Robert Montagu, Lord Mandeville (1649-1654), William Hammond (1655-1658), Banaster Maynard (1660-1663)* (14th Volume). London, UK: The Hakluyt Society.

Brewer, E., & Cunningham, K. (Eds.). (2010). *Integrating study abroad into the curriculum:*
Theory and practice across the disciplines. Sterling: VA: Stylus.

Brewer, E., & Leask, B. (2012). Internationalization of the curriculum. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 245-267). Thousand Oaks, CA: SAGE Publications, Inc.

Brickman, W.W. (1961). The meeting of East and West in educational history. *Comparative Education Review, 5*(2), 82-89.

Brophy, I.N. (1946). The luxury of anti-Negro prejudice. *Public Opinion Quarterly, 9*(4), 56-66. doi:10.1086/265762

Bryant, K. M., & Soria, K. M. (2015). College Students' Sexual Orientation, Gender Identity, and Participation in Study Abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad, 25*, 91-106.

Buckner, E., & Stein, S. (2019). What Counts as Internationalization? Deconstructing the Internationalization Imperative. *Journal of Studies in International Education*, Advance online publication. doi: 1028315319829878.

* Burton, S. (2013). Cultural and Cognitive Development in Short-Term Study Abroad: Illuminating the 360 Experience. (Doctoral dissertation). Michigan State University, East Lansing.

* Burrow, J. (2010). *Motivation and learning outcomes: A study of incoming exchange students at Queen's University* (Unpublished Master's thesis). Queen's University, Kingston.

Buzzard, J. (1993). *The Beaten Track: European Tourism, Literature, and the Ways to 'Culture', 1800-1918*. Oxford, UK: Clarendon Press.

Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence.* Clevedon, UK: Multilingual Matters.

* Caffrey, R. A., Neander, W., Markle, D., & Stewart, B. (2005). Improving the cultural competence of nursing students: Results of integrating cultural content in the curriculum and an international immersion experience. *The Journal of Nursing Education*, 44(5), 234-240.

Cai, W. W., & Sankaran, G. (2015). Promoting critical thinking through an interdisciplinary study abroad program. *Journal of International Students*, *5*(1), 38. Retrieved from https://digitalcommons.wcupa.edu/earmided_facpub/4

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental design for research*. Chicago, IL: Rand McNally.

Canadian Bureau of International Education [CBIE] (n.d.). Retrieved from https://cbie.ca/who-we-are/

Canadian University Survey Consortia. (2012). *2012 CUSC survey of graduating undergraduate students*. Retrieved from http://www.uregina.ca/cusc/publications/CUSC_2012_Graduating_Student_Survey_MasterReport.pdf

Capes, W. P. (1922). *The Modern City and Its Government*. New York, NY: Dutton

Card, N. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford.

Carlson, J., Burn, B., Useem, J., & Yachimovicz, D. (1990). *Study abroad: The experience of American undergraduates*. Westport, CT: Greenwood Press.

Carlson, J., & Widaman, K.F. (1988). The effects of study abroad during college on attitudes toward other cultures. *International Journal of Intercultural Relations, 12.* doi:10.1016/0147-1767(88)90003-X

Chieffo, L., and Griffiths, L. (2009). Here to stay: Increasing acceptance of short-term study abroad programs. (pp. 365–380). In R. Lewin (Ed.) *The handbook of practice and research in study abroad: Higher education and the quest for global citizenship*. New York, NY: Routledge.

Chieffo, L., & Griffiths, L. (2004). Large-scale assessment of student attitudes after a short-term study abroad program. *Frontiers: The interdisciplinary journal of study abroad, 10*, 165-177.

* Christopher Brooks, D. (2008). Learning tolerance revisited: A quasi-experimental replication. *Journal of Political Science Education*, *4*(3), 286-297. doi:10.1080/15512160802202664

Clarke, I., Flaherty, T. B., Wright, N. D., and McMillen, R. M. (2009). Student intercultural proficiency from study abroad programs. *Journal of Marketing Education, 31*(2), 173–181. doi:10.1177/0273475309335583

Coehlo, G. V. (1962). Personal growth and educational development through working and studying abroad. *Journal of Social Issues, 18*(1), 55-67. doi:10.1111/j.1540-4560.1962.tb02571.x

Commission on the Abraham Lincoln Study Abroad Fellowship Program (2005). *Global competence and national needs: One million Americans studying abroad.* Retrieved from https://www.nafsa.org/uploadedFiles/NAFSA_Home/Resource_Library_Assets/CCB/lincoln_commission_report(1).pdf

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field setting.* Boston, MA: Houghton Mifflin.

Cooper, H., & Hedges, L. V. (1994) (eds). *The Handbook of Research Synthesis*. New York, NY: Russell Sage Foundation.

Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods, 2*(4), 447-452. doi:10.1037/1082-989X.2.4.447

Craig, C. J., Zou, Y., & Curtis, G. (2018). The developing knowledge and identity of an Asian-American teacher: The influence of a China study abroad experience. *Learning, Culture and Social Interaction, 17*, 1-20. doi:10.1016/j.lcsi.2017.09.002

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college a meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272-295. doi:10.3102/0034654310362998

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cui, Q. (2013). *Global-mindedness and inter-cultural competence: A quantitative study of pre-service teachers* (Unpublished doctoral dissertation). Indiana State University, Terre Haute, Indiana.

Cushner, K., & Karim, A. U. (2004). Study abroad at the university level. *Handbook of intercultural training*, 289-308. doi:10.4135/9781452231129.n12

Cushner, K., & Mahon, J. (2002). Overseas student teaching: Affecting personal, professional, and global competencies in an age of globalization. *Journal of Studies in International Education, 6*(1), 44-58. doi:10.1177/1028315302006001004

Davis, F. J. (1963). Perspectives of Turkish students in the United States. *Sociology and Social Research*, 48, 47-57.

Davis, S. L., & Finney, S. J. (2006). A factor analytic study of the cross-cultural adaptability inventory. *Educational and Psychological Measurement, 66*(2), 318-330. doi:10.1177/0013164405278571

* Day-Vines, N. L. (1998). *Study abroad: an investigation of the impact of African diasporic travel on the psychosocial development of African American college sojourners (*Unpublished doctoral dissertation). North Carolina State University, Raleigh.

* DeLoach, S. B., Kurt, M., & Olitsky, N. H. (2015). Does content matter? Analyzing the change in global awareness between business- and nonbusiness-focused short-term study abroad courses. *Journal of Teaching in International Business*, *26*(1), 4-31. doi:10.1080/08975930.2014.929512

de Wit, H, Hunter, F., Coelen R. (2015). *Internationalization of Higher Education in Europe: future directions.* Brussels, Belgium: European Parliament.

Deardorff, D. K. (2004). *The identification and assessment of intercultural competence as a student outcome of internationalization at institutions of higher education in the United States* (Unpublished doctoral dissertation). North Carolina State University, Raleigh.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241-266. doi:10.1177/1028315306287002

Deardorff, D. K. (2009). Understanding the challenges of assessing global citizenship. In R. Lewin (Ed.), *The handbook of practice and research in study abroad: Higher education's quest for global citizenship* (pp. 346-364). New York: Routledge.

Deardorff, D. K. (2011). Assessing intercultural competence. *New Directions for Institutional Research*, *2011*(149), 65-79. doi:10.1002/ir.381

Deardorff, D. & Jones, E. (2012). Intercultural competence: an emerging focus in international higher education. In D. K. DeardorffH. d. Wit & J. D. Heyl *The SAGE handbook of international higher education* (pp. 283-304). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397.n16

Deardorff, D. K., Wit, H. D., Heyl, J. D., & Adams, T. (Eds.). (2012). *The SAGE handbook of international higher education*. Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397

Deng, F., & Tang, J. (2015). Introduce to the Non-symmetry of Word Derivation between 'Wenhua'and 'Culture'. *International Journal of Linguistics*, *3*(1), 145-150.

Denson, N. (2009). Do curricular and cocurricular diversity activities influence racial bias? A meta-analysis. *Review of Educational Research, 79*(2), 805-838. doi:10.3102/0034654309331551

Department of Foreign Affairs and International Trade [DFAIT]. (2014). *Canada's international education strategy: Harnessing our knowledge advantage to drive innovation and prosperity.* Ottawa, ON: Author.

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

Dhondt, P. (2008). A difficult balance between rhetoric and practice: Student mobility in Finland and other European countries from 1800 to 1930. In Byram, M., & F. Dervin (Eds.). 2008. *Student, staff and academic mobility in higher education*. Newcastle, UK: Cambridge Scholars Publishing.

Dillman, D. A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail and mixed-mode surveys: The tailored design method*, 3rd edition. New York, NY: John Wiley and Sons.

Dolby, N. (2004). Encountering an American self: Study abroad and national identity. *Comparative Education Review, 48*(2), 150–173. doi:10.1086/382620

Dolby, N. (2007). Reflections on nation: American undergraduates and education abroad. *Journal of Studies in International Education, 11*(2), 141–156. doi:10.1177/1028315306291944

Douglas, C., & Jones-Rikkers, C. G. (2001). Study abroad programs and American student world-mindedness: An empirical analysis. *Journal of Teaching in International Business, 13*(1), 55-66. doi:10.1300/J066v13n01_04

Doyle, D. (2009). Holistic assessment and the study abroad experience. *Frontiers: The Interdisciplinary Journal of Study Abroad, 18*, 143-156.

Dwyer, M. M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 151-164.

* Edwards, N. T. (2009). *The correlation between learner autonomy and cultural sensitivity in Japanese university students studying abroad (*Doctoral dissertation). Retrieved from Proquest Umi Dissertation Publishing.

* Engberg, M. E., & Jourian, T. J. (2015). Intercultural wonderment and study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad, 25*, 1-19.

Engberg, M. E., Jourian, T. J., & Davidson, L. M. (2016). The mediating role of intercultural wonderment: connecting programmatic components to global outcomes in study abroad. *Higher Education*, *71*(1), 21-37.

Engle, L., & Engle, J. (2003). Study abroad levels: Toward a classification of program types. *Frontiers: The interdisciplinary journal of study abroad, 9*(1), 1-20.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, 33*(5), 517. doi:10.1037/0003-066X.33.5.517.a

European Union. (2011). *Erasmus programme in 2011-12: The figures explained*. Retrieved from http://europa.eu/rapid/press-release_MEMO-13-647_en.htm

* Fabregas Janeiro, M. (2009). *Assessing changes in intercultural sensitivity in students exposed to intercultural experiences supported by the college of agricultural sciences and natural resources at Oklahoma State University using the intercultural development inventory* (Unpublished doctoral dissertation). Oklahoma State University, Still Water.

* Fairchild, S., Pillai, V., & Noble, C. (2006). The impact of a social work study abroad program in Australia on multicultural learning. *International Social Work, 49*(3), 390-401. doi:10.1177/0020872806063413

Fantini, A. E. (2006). *Exploring and assessing intercultural competence*. Retrieved from http://www.sit.edu/publications/docs/feil_research_report.pdf

Feldman, KA, & Newcomb, T.M. (1969). *The impact of college on students*. San Francisco, CA: Jossey-Bass.

* Fernandez, E. (2006). *Developing a global perspective during a study-term abroad* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage.

Fisher, R. A. (1925, July). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 22, No. 5, pp. 700-725). Cambridge, UK: Cambridge University Press.

Forum on Education Abroad (2011). *Education Abroad Glossary.* Retrieved from http://www.forumea.org/EducationAbroadGlossary2ndEdition2011.cfm

Forsey, M., Broomhall, S., & Davis, J. (2012). Broadening the mind? Australian student reflections on the experience of overseas study. *Journal of Studies in International Education, 16*(2), 128-139. doi:10.1177/1028315311407511

Forum on Education Abroad [FEA]. (2018). *Who we are.* Retrieved from https://forumea.org/about-us/who-we-are/

* Gaia, A. C. (2015). Short-term faculty-led study abroad programs enhance cultural exchange and self-awareness. *International Education Journal: Comparative Perspectives*, *14*(1), 21-31.

Gelman, A., & Weakliem, D. (2009). Of beauty, sex and power: Too little attention has been paid to the statistical challenges in estimating small effects. *American Scientist*, *97*(4), 310-316. Retrieved from https://www.jstor.org/stable/27859361

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 5*(10), 3-8. doi:10.3102/0013189X005010003

Gmelch, G. (1997). Crossing cultures: Student travel and personal development. *International Journal of Intercultural Relations*, *21*(4), 475–490. doi:10.1016/S0147-1767(97)00021-7

*Gingerich, K.E. (1998). *The impact of study abroad and didactic cross-cultural coursework experiences on the development of white racial consciousness and cultural sensitivity* (Unpublished doctoral dissertation). University of Kansas, Lawrence.

Golay, P. (2006). *The effects of study abroad on the development of global-mindedness among students enrolled in international programs at Florida State University* (Unpublished doctoral dissertation). Florida State University, Tallahassee.

Goldstein, S. B., & Kim, R. I. (2006). Predictors of U.S. college students' participation in study abroad programs: A longitudinal study. *International Journal of Intercultural Relations, 30,* 507–521. doi:10.1016/j.ijintrel.2005.10.001

Goodman, S. N. (1999). Toward evidence-based medical statistics: The p value fallacy. *Annals of Internal Medicine, 130*(12), 995-1004. doi:10.7326/0003-4819-130-12-199906150-00008.

Gore, J. E. (2005). *Dominant beliefs and alternative voices: Discourse, belief, and gender in American study abroad*. New York, NY: Routledge.

Government of Canada (2019). *Budget 2019.* Retrieved April 12, 2019 from https://budget.gc.ca/2019/docs/plan/chap-01-en.html#canada-s-new-international-education-strategy

Green, M., Marmolejo, F., & Egron-Polak, E. (2012). The internationalization of higher education: Future prospects. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 439-457). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397.n24

* Greenfield, E. A., Davis, R. T., & Fedor, J. P. (2012). The effect of international social work education: Study abroad versus on-campus courses. *Journal of Social Work Education, 48*(4), 739-761. doi: 10.5175/JSWE.2012.201100147

* Gullekson, N. L., Tucker, M. L., Coombs Jr, G., & Wright, S. B. (2011). Examining intercultural growth for business students in short-term study abroad programs:

Too good to be true? Journal of Teaching in International Business, 22(2), 91-106. doi:10.1080/08975930.2011.615672

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology*. (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Grünzweig, W., & Rinehart, N. (Eds.). (2002). *Rockin' in Red Square: Critical approaches to international education in the age of cyberculture*. Munster, Germany: Lit Verlag.

Gu, Q., Schweisfurth, M., & Day, C. (2010). Learning and growing in a 'foreign' context: Intercultural experiences of international students. *Compare: A Journal of Comparative and International Education, 40*(1), 7-23. doi:10.1080/03057920903115983

Guthrie, G. M., & Zektick, I. N. (1967). Predicting performance in the Peace Corps. *The Journal of Social Psychology*, *71*(1), 11-21. doi:10.1080/00224545.1967.9919761

Gullahorn, J. and Gullahorn, J. (1960). The role of the academic man as a cross-cultural mediator. *American Sociological Review, 25*(3), 414-417.

Gullahorn, J. T., & Gullahorn, J. E. (1963). An Extension of the U-Curve Hypothesis. *Journal of Social Issues, 19*(3), 33-47.doi:10.1111/j.1540-4560.1963.tb00447.x

* Gwynne, M. A. (1981). *The effects of study abroad on community college students* (Unpublished doctoral dissertation). Columbia University Teachers College, New York.

Hadis, B. F. (2005). Gauging the impact of study abroad: How to overcome the limitations of a single-cell design. *Assessment and Evaluation in Higher Education, 30*(1), 3-19. doi: 10.1080/0260293042003243869

Hall, E. (1959). *The silent language*. New York, NY: Doubleday.

Hamir, H. B. (2011). *Go abroad and graduate on-time: Study abroad participation, degree completion, and time-to-degree* (Doctoral dissertation). University of Nebraska, Lincoln.

Hammer, M. R. (1984). The effects of an intercultural communication workshop on participants' intercultural communication competence: An exploratory study. *Communication Quarterly, 32*(4), 252-262. doi:10.1080/01463378409369561

Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003). Measuring intercultural sensitivity: The intercultural development inventory. *International Journal of Intercultural Relations*, *27*(4), 421- 443. doi:10.1016/S0147-1767(03)00032-4

Hammer, M. R., Gudykunst, W. B., & Wiseman, R. L. (1979). Dimensions of intercultural effectiveness: An exploratory study. *International Journal of Intercultural Relations, 2*(4), 382-393. doi:10.1016/0147-1767(78)90036-6

* Hansen, R. A. (2010). Impact of study abroad on ethnocultural empathy and global-mindedness (Unpublished doctoral dissertation). Ball State University, Muncie.

Hanvey, R. G. (1982). An attainable global perspective. *Theory into Practice, 21*(3), 162-167. doi:10.1080/00405848209543001

Harrison, N. (2015). Practice, problems and power in 'internationalisation at home': Critical reflections on recent research evidence. *Teaching in Higher Education, 20*(4), 412-430. doi:10.1080/13562517.2015.1022147

* Harvey, T. A. (2013). *Facilitating intercultural development during study abroad: a case study of CIEE's seminar on living and learning abroad* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Hawthorne, L. (2012). Designer immigrants?: International students and two-step migration. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 417-437). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397.n23

Hazelkorn, E. (2015). Globalization, internationalization and rankings. *International Higher Education,* 8-10. doi:10.21427/D7TG9W

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational and Behavioral Statistics, 7*(2), 119-137. doi:10.3102/10769986007002119

Hett, E. J. (1993). *The development of an instrument to measure global-mindedness* (Unpublished doctoral dissertation). University of San Diego, San Diego.

Higgins, J., & Green, S. (Eds.) (2011) *Cochrane Handbook for Systematic Reviews of Interventions (*5.1). The Cochrane Collaboration.

*Hoff, J. G. (2005). *Students' perceptions of the culture learning process during the study abroad experience* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Hoffa, W. W. (2007). *A history of U.S. study abroad: Beginnings to 1965.* Carlisle, PA: Forum on Education Abroad.

Hoffa, W. W., & DePaul, S. C. (2010). *A history of U.S. study abroad: 1965–present.* Carlisle, PA: Forum on Education Abroad.

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of international business studies*, *14*(2), 75-89. doi:10.1057/palgrave.jibs.8490867

Hu, S., & Kuh, G. D. (2003). Diversity experiences and college student learning and personal development. *Journal of College Student Development*, *44*(3), 320-334. doi:10.1353/csd.2003.0026

Hudzik, J., & Stohl, M. (2012). Comprehensive and strategic internationalization of U.S. higher education. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 61-81). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452218397.n4

Hudzik, J. K. (2011). *Comprehensive internationalization*. Washington, D.C.: NAFSA

Hudzik, J. K. (2015). Integrating institutional policies and leadership for 21st century internationalization. *International Higher Education*, (83), 5-7. doi:10.6017/ihe.2015.83.9075

* Hughes, K. E. (2003). The effect of the International Business Institute study abroad program on cross-cultural adaptability among international business students (Unpublished doctoral dissertation). Capella University, Minneapolis, MN.

Hullihen, W. (1928). Present status of the junior year abroad. *The French Review*, *1*(2), 25-37. Retrieved from https://www.jstor.org/stable/380330

* Hunley, H. A. (2008). *The Impact of Mental Health on the Decision to Study Abroad and on Students' Functioning While Studying Abroad* (Unpublished doctoral dissertation). Loyola University, Chicago.

Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York, NY: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F., L. (20174). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.) Thousand Oaks, CA: SAGE.

Hurst, A. L. (2018). Class and Gender as predictors of study abroad participation among US liberal arts college students. *Studies in Higher Education*, 1-15.

* Hyndman III, W. T. (2009). *Cultural intelligence and self-efficacy as learning outcomes of study abroad* (Doctoral dissertation). Retrieved from ProQuest Information & Learning.

Indiana University. (1879). *Original ad for IU Summer Tramps, a faculty-led program established in 1879*. Retrieved from http://overseas.iu.edu/img/summertramps.jpg

Ingraham, E. C., & Peterson, D. L. (2004). Assessing the impact of study abroad on student learning at Michigan State University. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 83-100.

Institute of International Education (IIE). (2012). *Open doors 2012 report on international educational exchange.* Retrieved from http://opendoors.iienetwork.org/?p5150839.

Institute of International Education (IIE). (2016). *Open doors 2016 report on international educational exchange.* Retrieved from https://www.iie.org/en/Research-and-Insights/Open-Doors/Data/US-Study-Abroad/Destinations

Institute of International Education (IIE). (2017). *Open doors 2017 report on international educational exchange.* Retrieved from https://www.iie.org/en/Research-and-Insights/Open-Doors/Data/US-Study-Abroad/Destinations

Institute of International Education (IIE). (2018). *Why study abroad*. Retrieved from https://www.iie.org/en/Programs/Generation-Study-Abroad/About/Why-Study-Abroad

Intercultural Development Inventory (2017). *The roadmap to intercultural competence using the IDI.* Retrieved from https://idiinventory.com/

International Association of Universities. (2010). *Internationalization of higher education: Global trends, regional perspectives.* Retrieved from http://www.iau-aiu.net/internationalization/pdf/Key_results_2009.pdf

Isabelli-García, C., Bown, J., Plews, J. L., & Dewey, D. P. (2018). Language learning and study abroad. *Language Teaching*, *51*(4), 439-484. doi:10.1017/S026144481800023X

* Jackson, J. (2008). Globalization, internationalization, and short-term stays abroad. *International Journal of Intercultural Relations*, *32*(4), 349-358. doi:10.1016/j.ijintrel.2008.04.004

* Johnson, L. E., & Battalio, R. (2008). Expanding the boundaries of special education preservice teachers: The impact of a six-week special education study abroad program. *International Journal of Special Education, 23*(3), 90-100.

Jones, G. A. (2004). Ontario higher education reform, 1995-2003: From modest modifications to policy reform. *Canadian Journal of Higher Education*, *34*(3), 39-54.

* Kafka, E. P. (1969). *The effects of overseas study on world-mindedness and other selected variables of liberal arts students (*Unpublished doctoral dissertation). Michigan State University, East Lansing.

Kamal, A. A., & Maruyama, G. (1990). Cross-cultural contact and attitudes of Qatari students in the United States. *International Journal of Intercultural Relations, 14*(2), 123-134. doi:10.1016/0147-1767(90)90001-D

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319-342. doi:10.1111/j.1745-3984.2001.tb01130.x

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64). Westport, CT: American Council on Education/Praeger.

Keating, X. D., Guan, J., Piñero, J. C., & Bridges, D. M. (2005). A meta-analysis of college students' physical activity behaviors. *Journal of American college health, 54*(2), 116-126. doi:10.3200/JACH.54.2.116-126

* Keefe, M. C. (2008). *Short-term study abroad: Impact on the development of global competencies at a public college of art and design in the northeast* (Unpublished doctoral dissertation). Johnson & Wales University, Providence.

Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Boston, MA: Harvard University Press.

* Kehl, K. L. (2006). *Differences in self-efficacy and global mindedness between short-term and semester-long study abroad participants at selected Christian universities* (Unpublished doctoral dissertation). Baylor University, Waco, Texas.

Kelley, C., & Meyers, J. (1995). Cross-Cultural Adaptability Inventory. Minneapolis, MN: National Computer Systems.

Kelly, L. J. (1996). *Implementing Astin's IEO model in the study of student retention: A multivariate time dependent approach*. Paper presented at the Annual Forum of the Association for Institutional Research, Albuquerque, NM.

King, P. M., & Baxter Magolda, M. B. (2005). A developmental model of intercultural maturity. *Journal of College Student Development, 46*(6), 571-592. doi:10.1353/csd.2005.0060

Kinginger, C. (2013). Identity and language learning in study abroad. *Foreign Language Annals, 46*(3), 339-358. doi:10.1111/flan.12037

* 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.

* Kitsantas, A., and Meyers, J. (2001, March). *Studying Abroad. Does it Enhance College Student Cross-Cultural Awareness*? Paper presented at combined Annual Meeting of the San Diego State University and the U.S. Department of Education Centers for International Business Education and Research (CIBER 2001), San Diego, CA.

Knight, J. (2004). Internationalization remodeled: Definition, approaches, and rationales. *Journal of Studies in International Education, 8*(1), 5-31. doi:10.1177/1028315303260832

Knight, J. (2006). *Internationalization of higher education: New directions, new challenges. (2005 International Association of Universities Global Survey Report)*. Paris, France: International Association of Universities.

Knight, J. (2014). Is internationalisation of higher education having an identity crisis?. In A.M. Maldono (Ed.), *The forefront of international higher education* (pp. 75-87). Dordrecht, Netherlands: Springer.

Kronholz, J.F., & Osborn, D.S. (2016) The impact of study abroad experiences on vocational identity among college students. *Frontiers: The Interdisciplinary Journal of Study Abroad, 27,* 70-84.

Knight, J. (2012). Concepts, rationales, and interpretive frameworks in the internationalization of higher education. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education*. (pp. 27-43). Thousand Oaks, CA: SAGE. doi: 10.4135/9781452218397.n2

Kuh, G., O'Donnell, K.O., & Schneider, C.G. (2017). HIPs at ten. *Change: The magazine of higher learning, 49(*5), 8-16. doi:10.1080/00091383.2017.1366805

* Kurt, M. R., Olitsky, N. H., & Geis, P. (2013). Assessing global awareness over short-term study abroad sequence: A factor analysis. *Frontiers: The Interdisciplinary Journal of Study Abroad, XXIII,* 22-41.

Lee, J. J. (2015). Engaging international students. In J. Quaye & S.R. Harper (Eds.) *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (pp. 105-120). New York, NY: Routledge.

* Lemmons, K. K. (2013). *Short-term study abroad Programs: Where they came from, how they work, and why they often don't* (Unpublished doctoral dissertation). Texas A&M University, College Station, Texas.

Leung, K., Ang, S., & Tan, M.L. (2014). Intercultural Competence. *Annual Review of Organizational Psychology and Organizational Behaviour, 1.* 489-519. doi: 10.1146/annurev-orgpsych-031413-091229

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.

* Lombardi, M. (2011). *A study on students' intercultural learning through short-term study abroad programs* (Unpublished doctoral dissertation) Northeastern University, Boston.

* Luchesi, J. (2014). *Study Abroad as a Multifaceted Approach to Supporting College Sophomores: Creating Optimal Environments to Promote Intercultural Maturity* (Doctoral dissertation). Retrieved from ProQuest LLC.

Lukosius, V., & Festervand, T. A. (2013). Marketing Study Abroad Programs: A Student Recruitment Model. *American Journal of Business Education, 6*(5), 483-494.

Lyons, K., Hanley, J., Wearing, S., & Neil, J. (2012). Gap year volunteer tourism: Myths of global citizenship? *Annals of Tourism Research, 39*(1), 361-378. doi:10.1016/j.annals.2011.04.016

Lysgaard, S. (1955). Adjustment in a foreign society: Norwegian Fulbright grantees visiting the United States. *International Social Science Bulletin, 7*,45-51.

* Maharaja, G. G. (2009). *An" Island" Study Abroad Program and Its Impact on the Intercultural Sensitivity and Cross-Cultural Adaptability of Its Participants: Perspectives from a Research Intensive University* (Doctoral dissertation). Retrieved from Proquest. (815958894; ED512764)

Marginson, S., & van der Wende, M. (2009). The New Global Landscape of Nations and Institutions. In *Higher Education to 2030, Volume 2: Globalisation* (pp. 17-57). Paris: OECD.

* Mapp, S. C. (2012). Effect of short-term study abroad programs on students' cultural adaptability. *Journal of Social Work Education, 48*(4), 727-737. doi:10.5175/JSWE.2012.201100103

Massey, J., & Burrow, J. (2012). Coming to Canada to study factors that influence students decisions to participate in international exchange. *Journal of Student Affairs Research and Practice, 49*(1), 83-100. doi: 10.1515/jsarp-2012-6177

Mayhew, M.J., Pascarella, E.T., Bowman, N.A., Rockenbach, A.R., Seifert, T.A.D., Terenzini, P.T., & Wolniak, G.C. (2016). *How college affects students: 21st century evidence that higher education works* (Vol. 3). Somerset, NJ: John Wiley and Sons.

McCormick, A. C., & McClenney, K. (2012). Will these trees ever bear fruit?: A response to the special issue on student engagement. *The Review of Higher Education, 35*(2), 307-333.

McGill Office of the Provost and Vice-Principal (Academic) (n.d.). *International Strategic Framework*. Retrieved from https://www.mcgill.ca/international/strategic-framework

McKeown, J. S. (2009). *The first time effect: The impact of study abroad on college student intellectual development*. Albany, NY: State University of New York Press.

McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry* (6th ed.). New York, NY: Pearson.

* Medina-Lopez-Portillo, A. (2004). Intercultural learning assessment: The link between program duration and the development of intercultural sensitivity. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*, 179-199.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-748. doi:10.1002/j.2333-8504.1994.tb01618.x

Metcalfe, S. (2007). *Achieving cultural competency: Educational outcomes of study abroad experiences*. Retrieved from http://php.scripts.psu.edu/users/s/l/slc126/Eportfolio/Evidence/Literature%20Review.pdf

Meyers, C., & Arnold, S. (2016). Student expectations and reflections of a study away course experience to Washington, DC. Journal of Applied Communications, *100*(2), 86-100

Milstein, T. (2005). Transformation abroad: Sojourning and the perceived enhancement of self-efficacy. *International Journal of Intercultural Relations*, *29*(2), 217–239. doi:10.1016/j.ijintrel.2005.05.005

Ministry of Advanced Education and Skills Development (MAESD). (2018). *Published plans and annual reports 2017-2018.* Retrieved from https://www.ontario.ca/page/published-plans-and-annual-reports-2017-2018-ministry-advanced-education-and-skills-development

Mischel, W. (1965). Predicting the success of Peace Corps volunteers in Nigeria. *Journal of Personality and Social Pscyhology, 1*(5), 510-517. doi:10.1037/h0021915

Mitchell, K. (2012). Student mobility and European identity: Erasmus study as a civic experience?. *Journal of Contemporary European Research, 8(*4), 490-518.

Murray-García, J, & Tervalon, M. (2017). Rethinking intercultural competence: Cultural humility in internationalising higher education. In *Intercultural Competence in Higher Education* (pp. 19-31). Routledge.

NAFSA (2017). *Short-term grants, long-term impact: Creating institutional change in study abroad.* Retrieved from www.nafsa.org/publications

NAFSA (2018). *About NAFSA: Mission, vision, values.* Retrieved from https://www.nafsa.org/About_Us/About_NAFSA/Mission_and_Vision/Mission,_Vision,_Values/

* Nam, K. A. (2011). *Intercultural development in the short-term study abroad context: A comparative case study analysis of global seminars in Asia (Thailand and Laos) and in Europe (Netherlands)* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Nash, D. (1976). The personal consequences of a year of study abroad. *The Journal of Higher Education, 47*(2),191-203. doi:10.1080/00221546.1976.11774030

National Center for Education Statistics (2016). *Total fall enrollment in degree-granting postsecondary institutions*. Digest of Education Statistics. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_306.10.asp?current=yes

National Survey of Student Engagement (NSSE). (2018). *NSSE 2018 high-impact practices: US summary percentages by student characteristics.* Retrieved from http://nsse.indiana.edu/2018_institutional_report/pdf/HIPTables/HIP.pdf

Newton, P. & Shaw, S. (2014). *Validity in educational & psychological assessment*. London, UK: SAGE Publications Ltd.

* Ng, P., Goddard, T., Gribble, N., & Pickard, C. (2012). International placements increase the cultural sensitivity and competency of professional health students: A quantitative and qualitative study. *Journal of Physical Therapy Education,* 26(*1*), 61–68

* Nguyen, A. (2015). *Documentation and development on intercultural competence in short-term study abroad* (Unpublished doctoral dissertation). Texas Christian University, Fort Worth.

Nguyen, N.T., Biderman, M.D., & McNary, L.D. (2010). A validation study of the cross-cultural adaptability inventory. *International Journal of Training and Development, 14*(2), 112-129.

* Nichols, K. P. (2011). *Fostering intercultural competence through study abroad: A gender-based analysis of individual and program factors influencing development* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Niehaus, E., & Inkelas, K.K. (2016). Understanding STEM majors' intent to study abroad. *College Student Affairs Journal, 34(*1), 70-84. doi:10.1353/csj.2016.0004

Niehaus, E., & Crain, L. K. (2013). Act local or global?: Comparing student experiences in domestic and international service-learning programs. *Michigan Journal of Community Service Learning, 20*(1), 31-41.

Nowak, D. E., & Aloe, A. M. (2014). The prevalence of pathological gambling among college students: A meta-analytic synthesis, 2005–2013. *Journal of Gambling Studies, 30*(4), 819-843. doi:10.1007/s10899-013-9399-0

Nolan, R., & Hunter, F. (2012). Institutional strategies and international programs: Learning from experiences of change. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 131-147). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452218397.n8

Oberg, K. (1953). Indian tribes of northern Mato Grosso, Brazil. *Institute of Social Anthropology Publication.* Smithsonian Institution, Washington, D. C

* Ogden, A. C. (2010). *Education abroad and the making of global citizens: Assessing learning outcomes of course-embedded, faculty-led international programming* (Doctoral dissertation). Retrieved from ProQuest LLC. (3420254)

Olebe, M., & Koester, J. (1989). Exploring the cross-cultural equivalence of the behavioral assessment scale for intercultural communication. *International Journal of Intercultural Relations, 13*(3), 333-347. doi:10.1016/0147-1767(89)90016-3

* O'Reilly, A., Ryan, D., & Hickey, T. (2010). The psychological well-being and sociocultural adaptation of short-term international students in Ireland. *Journal of College Student Development*, 51(5), 584-598. doi: 10.1353/csd.2010.0011

Ordorika, I., & Lloyd, M. (2015). International rankings and the contest for university hegemony. *Journal of Education Policy, 30*(3), 385-405. doi:10.1080/02680939.2014.979247

Organization for Economic Cooperation and Development (OECD). (1996). *Internationalising the curriculum in higher education*. Paris, France: Author.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2)157-159. doi:10.3102/10769986008002157

* Pachmayer, A. (2014). *Enlightened Travelers? Cultural Attitudes, Cultural Competency and Study Abroad (*Unpublished doctoral dissertation). Arizona State University, Tempe.

* Paige, R. M., Cohen, A. D., & Shively, R. L. (2004). Assessing the impact of a strategies-based curriculum on language and culture learning abroad. *Frontiers: The interdisciplinary journal of study abroad, 10*, 253-276.

Paige, R. M, Fry, G. W., Stallman, E.M., Josić, J., & Jon, J.-E. (2009). Study abroad for global engagement: The long-term impact of mobility experiences. *Intercultural Education, 20*, 529–544. doi:10.1080/14675980903370847

Paige, R. M, Jacobs-Cassuto, M., Yershova, Y. A., & DeJaeghere, J. (2003). Assessing intercultural sensitivity: An empirical analysis of the Hammer and Bennett Intercultural Development Inventory. *International Journal of Intercultural Relations, 27*(4), 467-486. doi:10.1016/S0147-1767(03)00034-8

* Palmer, J. (2009). *Student acculturation, language preference, and L2 competence in study abroad programs in the Arabic-speaking world* (Unpublished doctoral dissertation). University of Arizona, Tucson.

Pascarella, E. T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research, 50*(4), 545-595. doi:10.3102/00346543050004545

Pascarella, E., & Terenzini, P. (1991). *How college affects students: Findings and insights from twelve years of research*. San Francisco, CA: Jossey-Bass.

Pascarella, E.T., & Terenzini, P.T. (2005). *How college affects students* (Vol. 2). San Francisco, CA: Jossey-Bass.

* Patterson, P. K. (2006). *Effect of study abroad on intercultural sensitivity.* (Unpublished doctoral dissertation). University of Missouri, Columbia.

* Pedersen, P. J. (2009). Teaching towards an ethnorelative worldview through psychology study abroad. *Intercultural Education, 20*(sup1), S73-S86. doi:10.1080/14675980903370896

* Pedersen, P. J. (2010). Assessing intercultural effectiveness outcomes in a year-long study abroad program. *International Journal of Intercultural Relations, 34*(1), 70-80. doi:10.1016/j.ijintrel.2009.09.003

Pendleton, C., Cochran, S., Kapadia, S., & Iyer, C. (2016). Understanding the gendered self: Implications from EI theory, the EI self, and the BEVI. In C. N. Shealy

(Ed.), *Making sense of beliefs and values: Theory, research, and practice* (pp. 261-301). New York, NY, US: Springer Publishing Co.

Perry, W. G. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York, NY: Holt, Rinehart, and Winston.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922-934. doi:10.1002/ejsp.504

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65-85.

Pitts, M.J. (2009). Identity and the role of expectations, stress, and talk in short-term student sojourner adjustment: An application of the integrative theory of communication and cross-cultural adaptation. *International Journal of Intercultural Relations, 33*(6), 450-462. doi:10.1016/j.ijintrel.2009.07.002

Pope, J. A., Sánchez, C. M., Lehnert, K., & Schmid, A. S. (2014). Why do Gen Y students study abroad? Individual growth and the intent to study abroad. *Journal of Teaching in International Business, 25*(2), 97-118. doi:10.1080/08975930.2014.896232

Porter, S. R. (2011). Do college student surveys have any validity? *Review of Higher Education, 35*(1), 45-76. doi:10.1353/rhe.2011.0034

Queen's University. (2011). *Queen's University Academic Plan 2011*. Retrieved from http://www.queensu.ca/saptf/wp-content/uploads/2011/10/Academic-Plan-for-SenatefinalNov22.pdf

Ramirez, E. (2002, November 15). *Strategies for Subject Matter Expert Review in Questionnaire Design*. Paper presented at the Questionnaire Design, Evaluation and Testing Conference. Charleston, SC.

Reilly, D., & Senders, S. (2009). Becoming the change we want to see: Critical study abroad for a tumultuous world. *Frontiers: The Interdisciplinary Journal of Study Abroad, 18*, 241-267.

* Rexeisen, R. J., & Al-Khatib, J. (2009). Assurance of learning and study abroad: A case study. *Journal of Teaching in International Business, 20*(3), 192-207. doi:10.1080/08975930903099077

* Rexeisen, R. J., Anderson, P. H., Lawton, L., & Hubbard, A. C. (2008). Study abroad and intercultural development: A longitudinal study. *Frontiers: The Interdisciplinary Journal of Study Abroad, 17*, 1-20.

* Rexisen, R. J. (2013). Study abroad and the boomerang effect: The end is only the beginning. *Frontiers: The Interdisciplinary Journal of Study Abroad, 22,* 166-181.

* Reza, A. (2015). *Fostering intercultural competence: Impacts of a multi-destination study abroad program* (Unpublished doctoral dissertation). Boston College, Newton.

Robinson, D. H. (2004). An interview with Gene V. Glass. *Educational Researcher, 33*(3), 26–30. doi: 10.3102/0013189X033003026

Rocconi, L. M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in higher education, 52*(2), 178-193. doi:10.1007/s11162-010-9190-3

Rocconi, L., & Gonyea, R. M. (May, 2015). *Contextualizing student engagement effect sizes: An empirical analysis*. Paper presented at the 2015 Association for Institutional Research Annual Forum, Denver, CO. Retrieved from http://nsse.indiana.edu/pdf/presentations/2015/AIR_2015_Rocconi_Gonyea_paper.pdf

* Roller, K. M. (2012). *Pre-Service Teachers and Study Abroad: A Reflective, Experiential Sojourn to Increase Intercultural Competence and Translate the Experience into Culturally Relevant Pedagogy* (Unpublished Doctoral Dissertation). University of California, Los Angeles.

Rosenthal, R (1979). The ‘‘file drawer problem’’ and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. doi:10.1037/0033-2909.86.3.638

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Hoboken, NJ: John Wiley & Sons.

Ruben, B. D. (1976). Assessing communication competency for intercultural adaptation. *Group & Organization Management, 1*(3), 334-354. doi: 10.1177/105960117600100308

Ruben, B. D. (1989). The study of cross-cultural competence: Traditions and contemporary issues. *International Journal of Intercultural Relations, 13*(3), 229-240. doi:10.1016/0147-1767(89)90011-4

Ruben, B. D., & Kealey, D. J. (1979). Behavioral assessment of communication competency and the prediction of cross-cultural adaptation. *International Journal of Intercultural Relations, 3*(1), 15-47. doi:10.1016/0147-1767(79)90045-2

* Rust, K. G., Forster, B., Niziolek, A., & Morris, C. M. (2013). Study abroad and intercultural coursework: Their effects on change in intercultural competence. *International Research and Review, 3(*1), 3-13.

* Rust, K. G., & Morris, C. M. (2013). Intercultural competency development in conjunction with a short term study away experience: Changes in undergraduate students. *China-USA Business Review*, *12*(8).

Ryan, M., and Twibell, R. (2000). Concerns, values, stress, coping, health, and educational outcomes of college students who studied abroad. *International Journal of Intercultural Relations, 24*(4), 409–435. doi:10.1016/S0147-1767(00)00014-6

Sá, C. M., & Sabzalieva, E. (2018). The politics of the great brain race: public policy and international student recruitment in Australia, Canada, England and the USA. *Higher Education*, *75*(2), 231-253.

* Sakurauchi, Y. H. (2014). *Teaching and Learning for Intercultural Sensitivity: A Cross-Cultural Examination of American Domestic Students and Japanese Exchange Student* (Doctoral dissertation). Retrieved from PDXScholar. (1643)

Salisbury, M. H. (2011). *The effect of study abroad on intercultural competence among undergraduate college students* (Unpublished doctoral dissertation). University of Iowa, Iowa.

Salisbury, M. H, Paulsen, M., & Pascarella, E. (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. doi: 10.1007/s11162-010- 9171-6

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

Salmi J. (2009). *The challenge of establishing world class universities.* Washington DC: The World Bank.

Sampson, D. L., & Smith, H. P. (1957). A scale to measure world-minded attitudes. *The Journal of Social Psychology*, *45*(1), 99-106. doi:10.1080/00224545.1957.9714290

* Savicki, V. (2010). Implications of Early Sociocultural Adaptation for Study Abroad Students. *Frontiers: The Interdisciplinary Journal of Study Abroad, 19*, 205-223.

* Savicki, V., Arrúe, C., & Binder, F. (2013). Language fluency and study abroad adaptation. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *XXII*, 37-57.

Savicki, V., & Cooley, E. (2011). American identity in study abroad students: contrasts, changes, correlates. *Journal of College Student Development, 52(*3), 339-349. doi:10.1353/csd.2011.0035

* Savicki, V., Downing-Burnette, R., Heller, L., Binder, F., & Suntinger, W. (2004). Contrasts, changes, and correlates in actual and potential intercultural adjustment. *International Journal of Intercultural Relations, 28*(3-4), 311-329. doi:10.1016/j.ijintrel.2004.06.001

Schrag, C. (1967). Elements of theoretical analysis in sociology. *Sociological theory: Inquiries and paradigms*, 220-253.

Seeber, M., Cattaneo, M., Huisman, J., & Paleari, S. (2016). Why do higher education institutions internationalize? An investigation of the multilevel determinants of internationalization rationales. *Higher education*, *72*(5), 685-702. doi:10.1007/s10734-015-9971-x

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(2), 173-199. doi:10.1017/S0272263105262027

Sell, D. K. (1983). Research on attitude change in U.S. students who participate in foreign study experiences: Past findings and suggestions for future research. *International Journal of Intercultural Relations*, *7*, 131-147. doi:10.1016/0147-1767(83)90017-2

S.601 - Senator Paul Simon Study Abroad Program Act of 2017 (2017). Retrieved October 25, 2018 from https://www.congress.gov/bill/115th-congress/senate-bill/601

* Shaheen, S. (2004). *The effect of pre-departure preparation on student intercultural development during study abroad programs* (Unpublished doctoral dissertation). Ohio State University, Columbus.

Shirley, S. W. (2006). *The gender gap in post-secondary study abroad: Understanding and marketing to male students* (Unpublished doctoral dissertation). University of North Dakota, Grand Forks.

Simpson. R. J. S., & Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, *2*(2288), 1243-1246. Retrieved from https://www.jstor.org/stable/20282622

Singleton, R. A., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Sinicrope, C., Norris, J., & Watanabe, Y. (2007). Understanding and assessing intercultural competence: A summary of theory, research, and practice (Technical report for the foreign language program evaluation project). *Second Language Studies, 26(1)*. 1-58. Retrieved from http://hdl.handle.net/10125/40689

* Smart, A. A. (2014). *Exploring the Impact of International Education Experiences on Change in Global Perspectives and American Identity Among College Students* (Unpublished Doctoral Dissertation). Northcentral University, Scottsdale.

* Smith, D. E., Smith, M. O., Robbins, K. R., Eash, N. S., & Walker, F. R. (2013). Traditionally Under-Represented Students' Perceptions of a Study Abroad Experience. *NACTA Journal*, 57(3a), 15-20. Retrieved from https://www.jstor.org/stable/nactajournal.57.3a.15

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*(9), 752-760. doi:10.1037/0003-066X.32.9.752

* Smith, T. K., & Moreno-Lopez, I. (2012). Outcomes of an interdisciplinary study abroad course: Learning Spanish and multicultural education concurrently. *Intercultural Education*, 23(4), 359-373. doi:10.1080/14675986.2012.724876

Sobania, N., & Braskamp, L. A. (2009). Study abroad or study away: It's not merely semantics. *Peer Review, 11*(4), 23-26

* Spenader, A. J., & Retka, P. (2015). The Role of Pedagogical Variables in Intercultural Development: A Study of Faculty Led Programs. *Frontiers: The Interdisciplinary Journal of Study Abroad, 25*, 20-36.

Spitzberg, B. H., & Cupach, W. R. (1984). *Interpersonal communication competence*. London, UK: Sage.

Spitzberg, B. H., & Cupach, W. R. (2007). The state of the art of stalking: Taking stock of the emerging literature. *Aggression and Violent Behavior, 12*(1), 64-86. doi:10.1016/j.avb.2006.05.001

Spitzberg, B.H., & Changnon, G. (2009). Conceptualizing intercultural competence. In D.K. Deadorff (Ed.), *The Sage handbook of intercultural competence* (pp. 2-52). Thousand Oaks, CA: SAGE Publications, Inc.

* Stallman, E. M. (2009). *Intercultural competence and racial awareness in study Abroad* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Stephan C.W., Stephan, W.G. (1992). Reducing intercultural anxiety through intercultural contact. *International Journal of Intercultural Relations, 16*, 89-106. doi:10.1016/0147-1767(92)90007-H

Stoof, A., Martens, R. L., Van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The boundary approach of competence: A constructivist aid for understanding and using the concept of competence. *Human Resource Development Review, 1*, 345-365. doi:10.1177/1534484302013005

Streitwieser, B., & Ogden, A.C. (2016). *International higher education's scholar-practitioners: bridging research and practice 2016.* Oxford, UK: Symposium Books Ltd.

* Stromberger, M. J. (2011). *Short-term study abroad and religious commitment: A look at what impacts global and multicultural awareness (*Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3435255)

Stroud, A. H. (2010). Who plans (not) to study abroad? An examination of U.S. student intent. *Journal of Studies in International Education, 20*(10), 1–18. doi:10.1177/1028315309357942

Sutton, S., Egginton, E., & Favela, R. (2012). Collaborating on the future: Strategic partnerships and linkages. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education*. (pp. 147-167). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452218397.n9

* Sutton, R. C., & Rubin, D. L. (2010). *Documenting the academic impact of study abroad:*
Final report of the GLOSSARI project. Paper presented at the NAFSA Annual Conference, Kansas City, MO.

Taskoh, A. K. (2014). *A Critical Policy Analysis of Internationalization in Postsecondary Education: An Ontario Case Study* (Doctoral dissertation). Retrieved from Western Electronic Thesis and Dissertation Repository. (1933)

Teichler, U. (2017). Internationalisation trends in higher education and the changing role of international student mobility. *Journal of international Mobility*, *1*, 177-216. doi:10.3917/jim.005.0179

* Teranishi, C. S., & Hannigan, T. P. (2008). Impact of study abroad on Latino/a college students' ethnic identity and cross-cultural adaptability. *Multicultural Learning and Teaching, 3*, 49-64. doi:10.2202%2F2161-2412.1029

Tervalon, M., & Murray-Garcia, J. (1998). Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education. *Journal of health care for the poor and underserved*, *9*(2), 117-125.

Tillman, M. (2012). Employer perspectives on international education. In D. Deardorff, H. Wit, J. Heyl, & T. Adams (Eds.), *The SAGE handbook of international higher education.* (pp. 191-207). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452218397.n11

Tippett, L.H. C. (1931). *The Methods of Statistics*. London, UK: Williams & Norgate.

Trilokekar, R. D., & Shubert, A. (2009). North of 49: Global citizenship à la Canadienne. In R. Lewin (Ed.), *The Handbook of practice and research in study abroad: Higher education and the quest for global citizenship (*pp. 191-212). New York, NY: Routledge.

Tsai, D.C. (1995). *Reliability of the Cross-Cultural Adaptability Inventory* (Unpublished doctoral dissertation). Rosemead School of Psychology, La Mirada, CA

* Tucker, M. L., Gullekson, N. L., & Edmond-Kiger, C. (2014). Accounting for EI: Does emotional intelligence predict greater intercultural growth? *Journal of International Business and Cultural Studies, 8*. 1-11.

Twombly, S. B., Salisbury, M. H., Tumanut, S. D., & Klute, P. (2012). Study abroad in a new global century: Renewing the promise, refining the purpose. *ASHE Higher Education Report, 38*(4), 1-152.

United States – China Strong (2018). *Initiatives:1 million strong*. Retrieved from http://100kstrong.org/initiatives/

United States Department of Education (2012). *Succeeding Globally Through International Education and Engagement U.S. Department of Education International Strategy 2012–16.* Retrieved from https://www2.ed.gov/about/inits/ed/internationaled/international-strategy-2012-16.pdf

United States State Department (2009, June 27). *100,000 Strong Educational Exchange Initiatives.* Retrieved from http://www.state.gov/100k/index.htm

University of British Columbia. (2011). *UBC international strategic plan.* Retrieved from https://research.ubc.ca/sites/research.ubc.ca/files/vpri/UBC-intl-strat-plan-2011.pdf

University of British Columbia Student Services (2018). *International experiences.* Retrieved from https://students.ubc.ca/career/international-experiences

University of Western Ontario. (2009). *The University of Western Ontario strategic plan for internationalization, 2009-2012.* Retrieved from http://www.uwo.ca/univsec/board/minutes/2009/r0906sen_ann1.pdf

University of Western Ontario. (2018). *Measuring our progress: Paying attention to metrics that matter.* Retrieved from https://president.uwo.ca/strategic_planning/progress.html

University of Toronto (2017a). *Global Engagement: U of T in the World. Office of the Vice-President, International.* Retrieved from https://global.utoronto.ca/wp-content/uploads/2015/08/U-of-T-International-Strategic-Plan.pdf

University of Toronto (2017b). *Results of the National Survey of Student Engagement 2017*. Office of the Provost. Retrieved from http://www.viceprovoststudents.utoronto.ca/Assets/Students+Digital+Assets/NSSE2017.pdf

University of Toronto Mississauga (2018). *UTM Across Canada*. Retrieved from https://www.utm.utoronto.ca/international/outbound-opportunities/utm-across-canada

Valdez, G. (2015). US Higher Education Classroom Experiences of Undergraduate Chinese International Students. *Journal of International Students*, *5*(2), 188-200.

Vande Berg, M., Connor-Linton, J., & Paige. R. M. (2009). The Georgetown consortium study: Intervening in student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad, 18*, 1-75.

Varela, O. E. (2017). Learning outcomes of study-abroad programs: A meta-analysis. *Academy of Management Learning & Education*, *16*(4), 531-561.

Varela, O. E., & Gatlin-Watts, R. (2014). The development of the global manager: An empirical study on the role of academic international sojourns. *Academy of Management Learning & Education*, *13*(2), 187-207.

Vella-Zarb, R. A., & Elgar, F. J. (2009). The 'freshman 5': a meta-analysis of weight gain in the freshman year of college. *Journal of American College Health*, *58*(2), 161-166.

Vogt, P. (2006). *Quantitative research methods for professionals*. Upper Saddle River, NJ: Pearson.

Ward, C., Okura, Y., Kennedy, A., & Kojima, T. (1998). The U-curve on trial: A longitudinal study of psychological and sociocultural adjustment during cross-cultural transition. *International Journal of Intercultural Relations*, *22*(3), 277-291. doi:10.1016/S0147-1767(98)00008-X

Warmbrod, J. R. (1986, December). *The theoretical/conceptual framework: What is its relevance to conclusions and recommendations.* Paper presented at Annual meeting of the American Educational Research Association, Dallas, TX.

Wahyudi, R. (2016). Intercultural competence: Multi-dynamic, intersubjective, critical and interdisciplinary approaches. In *Intercultural Competence in Education* (pp. 143-166). Palgrave Macmillan, London.

Williams, R.M. Jr. (1947). *The reduction of intergroup tensions: a survey of research on problems of ethnic, racial, and religious group relations.* New York, NY: Social Science Research Council.

Williams, R. (2006). The Analysis of Culture. In J. Storey (Ed.). *Cultural theory and popular culture: A reader*. (pp. 32-40). University of Georgia Press.

* Williams, T. R. (2002). *Exploring the impact of study abroad on students' intercultural communication skills: Adaptability and sensitivity* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 1410170)

Williams, T. R. (2005). Exploring the impact of study abroad on students' intercultural communication skills: Adaptability and sensitivity. J*ournal of Studies in International Education, 9*(4), 356-371. doi:10.1177/1028315305277681

Willis, G. (1999). *Cognitive interviewing: A how to guide.* Course presented at the 1999 Meeting of the American Statistical Association. Retrieved from http://fog.its.uiowa.edu/~c07b209/interview.pdf

Wilson, D. B. (2009). Systematic coding for research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 159 –176). New York, NY: Russell Sage Foundation.

Woolf, M. (2007). Impossible things before breakfast: Myths in education abroad. *Journal of Studies in International Education, 11*(3-4), 496-509. doi:10.1177/1028315307304186

World Bank (2019). Databank. *Educational attainment, at least completed post-secondary, population 25+, total (%) (cumulative).* Retrieved April 14, 2019 from https://data.worldbank.org/indicator/SE.SEC.CUAT.PO.ZS?locations=CA

* Wortman, T. I. (2002). *Psychosocial effects of studying abroad: Openness to diversity* (Unpublished doctoral dissertation). Pennsylvania State University, Pennsylvania.

Yang, J. S. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *Canadian Modern Language Review, 72*(1), 66-94. doi:10.3138/cmlr.2344

* Zarges, B. P. (2016). *The effect of short-term international immersion experiences on the development of intercultural competence in students at a faith-based*

university (Unpublished doctoral dissertation). Indiana State University, Terre Haute.

Zemach-Bersin, T. (2007). Global citizenship & study abroad: It's all about U.S. critical literacy. *Theories and Practices, 1*(2), 16–28.

Zemach-Bersin, T. (2008). American students abroad can't be 'global citizens.'. Chronicle of Higher Education, 54(26), A34.

Zemach-Bersin, T. (2009). Selling the world: Study abroad marketing and the privatization of global citizenship. In R. Lewin (Ed.), *The handbook of practice and research in study abroad: Higher education and the quest for global citizenship* (pp. 303–320). New York, NY: Routledge.

Zemach-Bersin, T. (2012). Entitled to the world: The rhetoric of US global citizenship education and study abroad. In V. de Oliveira Andreotti & L.M.T.M de Souza (Eds.), *Postcolonial perspectives on global citizenship education* (pp. 99-116). New York, NY: Routledge.

* Zhai, L. (2000). *The influence of study abroad programs on college student development in the College of Food, Agricultural, and Environmental Sciences at the Ohio State University* (Unpublished doctoral dissertation). Ohio State University, Columbus.

* Zielinski, B. A. Z. (2007). *Study abroad length of program influence on cross-cultural adaptability* (Unpublished master's thesis). Virginia Polytechnic Institute and State University, Blacksburg, VA