Research Article

Frontiers: The Interdisciplinary Journal of Study Abroad

Volume 36, Issue 3, pp. 115-139
DOI: 10.36366/frontiers.v36i3.769
www.frontiersjournal.org

Assessing the Effects of Semester-Length Study Abroad with the Miville-Guzman Universality-Diversity Scale

Zachary Nowak[1], Jacopo Bernardini[1], Paola Cascinelli[2], Amanda Fontana[3]

Abstract

There is a vibrant literature dedicated to evaluating the effects of international education experiences on the students who participate in them. While much discourse centers on the development of "(inter) cultural competencies," research on assessment suggests that these competencies develop over a longer period of time than most abroad programs last, hindering educators' ability to attribute their development to these programs. In this paper, we present the results of our research from two semesters of our program. We collected quantitative and qualitative data relative to students' Universal-Diverse Orientation (UDO). The quantitative data was gathered through the short version of the Miville-Guzman Universality-Diversity Scale (M-GUDS-S). For qualitative data, we relied on open-ended questions and reflective journaling assignments. We found statistically significant quantitative changes in both semesters. The qualitative data reflected and helped explain these changes. We discuss the implications of these results for future research.

Abstract in Italian

Esiste una vivace letteratura dedicata alla valutazione degli effetti delle esperienze educative internazionali sugli studenti che vi partecipano. Mentre gran parte del dibattito è incentrato sullo sviluppo di "competenze (inter)culturali", la ricerca valutativa suggerisce che queste competenze si sviluppino in un periodo di tempo più lungo rispetto alla durata della maggior parte dei programmi all'estero,

1 The Umbra Institute: Study Abroad, Perugia, Italy
2 Arcadia University, Glenside, PA, United States of America
3 ISI Abroad, Orlando, FL, United States of America

Corresponding author: Zachary Nowak, znowak@umbra.org

Date of Acceptance: June 24th, 2024

ostacolando la capacità degli educatori di attribuire a questi programmi i cambiamenti misurati. In questo articolo presentiamo i risultati della nostra ricerca raccolti durante due dei semestri del nostro programma. Abbiamo raccolto dati quantitativi e qualitativi relativi all'Orientamento Universale-Diverso (UDO) degli studenti. I dati quantitativi sono stati raccolti attraverso la versione breve della scala Miville-Guzman Universality-Diversity (M-GUDS-S). Per i dati qualitativi, abbiamo fato riferimento ai compiti assegnati agli studenti che prevedevano un diario e domande a risposta aperta. Abbiamo riscontrato cambiamenti quantitativi statisticamente significativi in entrambi i semestri. I dati qualitativi hanno confermato tali risultati e hanno aiutato a spiegare i cambiamenti. Le implicazioni di questi risultati per future ricerche saranno discusse.

Keywords

Assessment; (cross-)cultural competencies; Miville-Guzman Universality-Diversity Scale (M-GUDS); UDO

1. Introduction

Just prior to the COVID-19 pandemic, Elspeth Jones noted the twin factors for integrating study abroad more intentionally into the higher education system. One factor was the pull: some sort of intercultural competence has become valued if not necessary for many jobs in an increasingly globalized and often digital marketplace. In parallel with the rising value of global experience are a host of competitive push factors: "changes in global student flows, international branch campuses, and growing complexity in cross-border activity, along with the rising influence of institutional rankings, all provide economic impetus and reputational consequences of success or failure" (Jones, 2018, P. xv).

The COVID-19 pandemic and the post-pandemic era have put even more pressure on international education professionals to prove the value of the programs they design and administer. The short-term consequences were obvious, as nearly all US students studying abroad were sent back to the US when the pandemic was declared in March 2020. Ninety-three percent of the planned summer study programs for 2020 were canceled, as were two thirds of Fall 2020 programs (Rogers, 2020, p. 4). In addition to US-based students studying abroad, the pandemic also had deleterious effects on international students from other countries either studying abroad for a semester or pursuing their degrees outside their country of residence (Ahlburg, 2020; Bilecen, 2020). Many US institutions subsequently either completely cut their international education offices or at least cut them back. Instructors who were used to organizing annual J-term or summer faculty-led programs now must not only restart interrupted promotional efforts, but also justify their programs' educational value and lack of significant risk in

ways they had not had to do prior to March 2020 (Raj Kumar et al. 2022). As the leaders of The GLOSSARI Project wrote a decade and a half ago, "the field of international education is moving forward to confront the challenges of data-driven, evidentiary-based articulation of the values gained from study abroad" (Sutton & Rubin, 2004). This move, however, has been slow as what we as international education professionals are attempting to measure is often hard to quantify.

These counterposed forces—institutional cost-cutting and increased scrutiny of objectives against the increasing importance of a less provincial education for career success—mean that the need for accurate assessment of international programs has become even more important. As Savicki and Brewer highlighted in the introduction to their volume on the assessment of these programs, the fundamental question is "What does study abroad achieve?" Disaggregating that question leads to a series of others: Who will assess what? What should be assessed? To what end? When, where, and how? And with "tests or texts," i.e., quantitative or qualitative measures (Savicki & Brewer, 2015, pp. 1–4)? The researchers in this present study recognized, as we dove into the literature on assessment, the limitations in our program's evaluation process, in the sense of a "systematic collection, analysis, and use of information collected about courses and programs for the purpose of improving courses and programs" (Deardorff & van Gaalen, 2021, p. 148). There is considerable evidence that international experiences, especially "high impact" ones, have a broad range of significant outcomes (see especially Braskamp et al., 2009; Stebleton et al., 2013). Our goal was to see whether our specific program actually had a (statistically) significant impact on students. Our academic program had learning outcome assessment as our courses use backwards design; we assess our instructors using course evaluations. We had long had a final program evaluation, but that fell squarely within the realm of what has been called the "customer satisfaction" model of assessment in study abroad (Engle, 2013; Vande Berg et al., 2012).

In this paper we offer some of our responses to these questions. In contrast to studies designed and carried out by international education offices with students returning from a variety of destinations, we have attempted to assess the expectations of and effects on students who completed a ten-week course of study at our institution in the academic year 2021-2022. We gave them a short pre- and post-test battery of questions that included the short version of the Miville-Guzman Universality-Diversity Scale (M-GUDS-S), several open-ended questions on their expectations of their experience abroad, the city of Perugia, and their time at The Umbra Institute, and finally several questions on the obligatory Italian language class which they would be taking while studying with us. In addition to this survey, which collected both quantitative and qualitative data, we also asked

the students to write (in English) short reflections in a course journal as part of their Italian course. Finally, we conducted a focus group. Our institutional goals were to assess the importance of language instruction to our students and what sort of equivalencies (if any) students would receive for that language instruction at their home institutions; to better understand students' expectations of the experience (before) and whether those expectations were satisfied (after); and, most importantly, to evaluate whether their ten weeks studying with us and living in Perugia had had a measurable impact on them, particularly on their global thinking and cultural engagement. This is particularly important because as Yakunina et al. (2012) have shown, "whereas personality traits are generally seen as fixed and resistant to change, openness to diversity may be a multicultural attitude that may be more amenable to development or modification" (p. 538). In other words, international education can have a (long-term) impact on students' ability to embrace people, ideas, and cultures different from their own.

In what follows, we first review the relevant literature on the assessment of study abroad, focusing on the debates about what can be assessed, the appropriate measures, and the results of recent assessments. We then explain our institutional objectives for undertaking this research and justify our methodological decisions in the context of what we could feasibly investigate. Finally, we present the results we obtained, then both analyze them and draw conclusions for future research.

2. Literature on Assessment

There are a wide variety of indices available for assessing conceptual constructs in which international education professionals are interested. Some of these instruments are used more broadly in other educational contexts and have been repurposed for education abroad. One such instrument is the Beliefs, Events, and Values Inventory (BEVI), an instrument used for more than three decades "to understand and facilitate processes and outcomes of learning, growth, and transformation" (BEVI, 2023; see also Shealy, 2016). The BEVI was used by the Forum on Education Abroad for a multi-year, multi-site assessment published in 2015 (Wandschneider et al., 2015) but has also been used on short-term programs (Grant et al., 2021) and even, most recently, online study abroad programs (Seo & Konishi, 2023). Similarly, the Global Perspective Inventory (GPI), a 72-question scale currently managed by Iowa State University and used by almost two hundred colleges, universities, and educational organizations to assess assesses global learning and development in various contexts (Iowa State University of Science and Technology, 2023). It has been used to show significant impacts of traditional education abroad (semester) programs, as well as the effect of service learning components in those programs (Engberg, 2013). The GPI has also been used to find pre- and post-test differences with short-term faculty-led programs

(Gaia, 2015), though sometimes the results on this sort of program have been counterintuitive (Whatley et al., 2021).

These are not the only indices that have been used for assessment of student growth in education abroad contexts. Other principal measures used in the past include the Intercultural Development Inventory (IDI) (Anderson et al., 2006; Fitzgerald et al., 2018; Nevado Llopis & Sierra Huedo, 2022) and the Intercultural Sensitivity Scale (ISI) (Hammer, 2012). Add to that a series of other measures—Global Competence Aptitude Assessment (GCAA), the Intercultural Effectiveness Scale, the Global Mindedness Scale, and the Intercultural Adjustment Potential Scale, Consumer Ethnocentric Tendencies Scale (CETSCALE), and the Cross Cultural Adaptability Inventory (CCAI) (for a review, see Schenker, 2019)— as well as ad hoc instruments created by the researchers themselves (for example Sutton & Rubin, 2004; Norris & Dwyer, 2005). While none of these indices can provide the whole picture of learning while studying abroad, each can contribute something to knowledge about education abroad's effect on students.

Despite their value in giving education abroad professionals programmatic feedback, previous studies demonstrating the beneficial effects of study abroad on students sometimes have shared methodological flaws: lack of a clearly defined conceptual model about what is changing, small samples from a single institution, selection bias, the lack of measuring change over time, not accounting for possibly confounding variables (Salisbury et al., 2013), or simply a cost that is not in within the budgetary constraints of programs like ours. When planning our study, we made a decision about which quantitative measure to use based on Bowman's three characteristics of good tools with which to assess student learning and growth: the indices used must be *valid*, *reliable*, and *affordable* (Bowman, 2010). These characteristics—especially the third —were central to our decision-making process and eliminated from consideration a number of the above indices. The BEVI has both quantitative and qualitative items, takes 30 minutes to administer by a (necessarily) trained administrator, and—for our organization of our (small) size—costs USD$3,000 annually. For the IDI, an organization has to pay over USD$2,000 to have a staff member trained to administer the test, plus USD$13 per student per test. Some other measures we decided were either too long, were not modifiable for our context, or did not measure what we ultimately were interested in measuring.

For this reason, we decided to use the short version of the Miville-Guzman Universality-Diversity Scale (M-GUDS) assessment. The M-GUDS was introduced by Marie Miville and her collaborators in 1999 and has proved reliable at assessing a construct that the authors called UDO: "an attitude of awareness of both the similarities and differences among people" (Miville et al., 1999, p. 291). This measure then assesses the extent to which people both recognize the

differences that separate them from others (sex, ethnicity, or–most importantly for us–nationality) but also recognize some level of kinship with people who differ from them along these axes. The UDO construct, as measured by the M-GUDS index, correlates positively with attitudes toward diversity of other people, with empathy, healthy narcissism, open-mindedness, and feminism, as well as negatively with homophobia and dogmatism (Fuertes et al., 2000; Miville et al., 1999; Yakunina et al., 2012).

The original scale had 45 questions, divided into three fifteen-question subscales: Diversity of Contact, Relativist Appreciation, and Comfort with Differences. Importantly, M-GUDS-S scores do not correlate with the social desirability scale, meaning that the answers are not simply what the respondents think the researchers want to hear; in addition, the researchers found the scale to be reliable, with α = .93 and a test-retest reliability of .94 (Miville et al., 1999). A later study by one of Miville's collaborators proposed a shorter version of the index, one which had fifteen questions (albeit still divided into three subscales). We used this shorter form of the M-GUDS scale (M-GUDS-S), which the original scale's developers determined to be as reliable and valid as the original scale while being only fifteen questions (Fuertes et al., 2000). In addition to the quantitative instrument, we also collected qualitative data. Following Stevens and Cooper's (2009) guidelines for effective course journaling, a colleague created a reflection journal that each student wrote in as part of their mandatory Italian course (this journal was called *My Semester at The Umbra Institute: Language, Culture, & Reflection* and was created by Elgin K. Eckert).

3. Overview and Objectives

This section provides more details on our institutional objectives for the present research, the logistical-didactic structure of our institution as well as the student body, the M-GUDS-S scale and other data we gathered, and the statistical and qualitative analyses we carried after data collection was complete. Some additional methodological notes can be found in Appendix B.

Our purpose in this research, carried out in two separate data-gathering periods (the two semesters of the 2021-2022 academic year) was to try to look for changes in students' openness to diversity and universality that could be a result of their semester abroad. This research was intended to see if the experience at our Institute did indeed have a beneficial effect on students and will serve as the basis of future research, which we intend to carry out to disaggregate which parts of the study abroad experience are the most impactful (and therefore which we should promote more). We were mindful of Deardorrff's comment, that "a multi-method approach of both direct and indirect methods is key in providing a more

comprehensive picture of student learning, especially with the complex outcomes found in international education" (Deardorff, 2015, p. 20).

3.1. The Institution and the Student Body

The Umbra Institute is an independent American study abroad program in the central Italian city of Perugia. Perugia has 165,000 residents as well as roughly 28,000 Italian and international students studying at its two main Italian degree-granting universities. The Umbra Institute has nearly forty partners from whom it receives students: in the two semesters most of roughly one hundred students take an obligatory Italian class as well as three or four courses in English, across an array of disciplines. Five percent of students take courses directly in Italian at the two Italian universities. Community engaged learning—whether through community-focused courses, service-learning components, or volunteer opportunities—has been at the center of the pedagogical approach since The Umbra Institute was founded in 1999. While for linguistic reasons—the fact that 90% of the students are native speakers of English, and must only take one course in Italian—The Umbra Institute is closest to Engle and Engle's "Level Three: Cross-Cultural Contact Program" category, for the residential and community-engaged learning aspects, it more closely resembles a "Level Five: Cross-Cultural Immersion Program" (Engle & Engle, 2003).

The wide range of educational institutions sending to our program avoid the weaknesses of assessing a relatively homogeneous sample from any one institution: the students were from private colleges and universities like Harvard, Bryn Mawr, and Clark as well as large public institutions such as Penn State University and the University of Connecticut; in Fall 2021 thirty-one institutions sent students, whereas in Spring 2022 there were students from thirty-eight US colleges and universities.

In terms of demographic data, during the Fall 2021 semester, a total of 52 students were enrolled in our programs, with 39 female and 13 male students. Regarding gender identification, 37 students identified as female, 11 as male, and four as neither. The majority of students (76%) fell within the age group of 19 to 21, while the remaining 24% were between 22 and 25. Moving to the Spring 2022 semester, we had a total of 94 students, comprising 75 females and 19 males. In terms of gender identification, 70 students identified as female, 17 as male, and seven as neither. Again, a significant proportion (83%) of the student body belonged to the age group between 19 and 21, with the remaining 17% falling between 22 and 25. It is noteworthy that these figures are meticulously reflected in our study, where all four samples (respondents of the initial and final surveys of both longitudinal studies) closely mirror the reference population. Any differences observed are statistically insignificant.

Furthermore, as outlined in Appendix B, we systematically examined the possible correlation between these demographic data and all items on the scale. Our thorough analysis did not reveal any statistically significant differences that would warrant specific mention. This comprehensive examination ensures that the demographic factors we have considered do not appear to influence the outcomes measured by the scale.

This varied sample avoided some of the major methodological flaws that have weakened previous studies: the characteristics of the institutions were different as were therefore at least some parts of students' precollege characteristics and their experiences within their individual colleges.

3.2. The M-GUDS-S Scale

The M-GUDS scales (both long and short versions) have been used frequently in research on young adults. Some authors have used it to assess possible correlations between academic performance in high school and UDO scores (Singley & Sedlacek, 2004). Other authors assessed the effect of a year-long residential civic learning community experience on undergraduates' scores (Longerbeam & Sedlacek, 2006). Overall, the M-GUDS index is a free, easy-to-use, and yet empirically validated means of assessing individuals' openness to diversity (Fuertes et al., 2000; Miville et al., 1999; Singley & Sedlacek, 2004). In addition to this quantitative index, we also used several sources of qualitative, open-ended data, given that "qualitative evaluations solicit judgment or conclusions about the value or merit of whatever performance, places, or events are targeted for review" (Engle, 2013, p. 115).

To carry out the quantitative part of this research, we used a slightly modified version of the M-GUDS-S scale (see Appendix A). We used the word "they" for greater gender inclusivity and replaced the word "race" with "culture" in several questions. While the pronoun choice perhaps speaks for itself, we think it appropriate to explain our choice about the word race. Certainly in Europe but also in many other parts of the world, the word race has been abandoned in scientific research given its negative connotations. As Klein and Ravenda (2016) state, using the word race risks confirming and legitimizing race as a classificatory category of human difference (p. 39; see also Cozien, 2015). There was a methodological risk of making this change. Students who have grown up in or at least studied in the U.S.—students for whom "race" *is* a meaningful category — being confronted with a battery of questions in which that word is not present, may have different responses precisely because of the absence of the word "race." We ultimately decided, though, that the change was worth the possible effect on the data; we plan to describe our modifications to the M-GUDS-S in a future paper, but those interested in our version can contact us in the meantime. In addition to

the fifteen M-GUDS-S questions, we asked three open-ended questions both semesters as well as adding several demographic questions for the second pre-test in Spring 2022.

3.3. Data Collection Methods

Our study used data from two semesters of students: Fall 2021 (n = 51) and Spring 2022 (n = 91 for the pretest, n = 81 for the post-test), collected using a Google form. In both cohorts, the response rate was high (approximately 98% in Fall and 92% in Spring) as the pre- and post-test surveys were an obligatory part of the first orientation meeting and the final community engagement presentations event. We did not analyze the data in terms of the three subscales of the M-GUDS-S as we had made the modifications described above in the wording and were not sure the subscales still were internally valid. In addition to the quantitative data from the M-GUDS-S scale, we also gathered qualitative data from three open-ended questions on the survey. Other qualitative data came from student reflection journals and a single focus group with seven students; the goal of the focus group was purely exploratory and helpful ultimately for the drafting of the open-ended questions and for giving us lines of analysis. For the quantitative data, we employed several statistical techniques using IBM's SPSS Statistics software for data mining and analysis. Firstly, we conducted descriptive statistics, including frequencies and descriptives. Frequencies were utilized to provide a count of occurrences for each variable, while descriptives, incorporating standard deviation, were employed to analyze the dispersion and variability within the data.

To further explore the relationships between multiple variables, particularly demographic variables, and scale items, we employed crosstabulation utilizing percentage of both rows and columns to present a detailed analysis of the relationships between demographic variables and scale items. Z-test was also employed to compare column proportions. This statistical technique was instrumental in assessing whether the differences observed in column proportions were statistically significant. For the qualitative data, we did a content analysis after aggregating the responses to each question.

4. Results

4.1. Quantitative Results: The numeric Snapshots

Despite their confidence in the transformative nature of international education in general and in the intentionally designed semester-long experience at The Umbra Institute, the authors of the present article had anticipated two scenarios for changes in the scores of the student participants. In one scenario, there would be no change in the mean scores for the fifteen different questions in

the index, but there would be statistically significant changes in the standard deviation. In other words, we anticipated that while we perhaps would be unable to show an increase in the mean scores, the range of scores would get closer to the mean (already high) score. The other situation we expected was that we would find neither a statistically significant increase in means nor a statistically significant change in the standard deviation. As the research was undertaken to assess and improve our program, we recognized that even the lack of a significant change could be diagnostic. For example, Grant et al. (2021) used the BEVI (Beliefs, Events, and Values Inventory) for their assessment of a very short-term (9-day) program and found no statistically significant changes; however, they argued that "synthesizing formative and summative assessment practices [can] suggest potential curriculum changes and additional resources to support student needs" (p. 129).

As Grant et al. (2021, p. 140) also point out, the students who choose to participate in international study—especially those who choose lesser-known destinations like our city, and especially in more specialized curricular tracks like those our program offers, we would add— likely tend to be those students who are already more culturally competent and (more to the point for this study) open to diversity. In addition to this survivor bias in the data, we expected that these two particular semesters, from September to December 2021 and February to April 2022, would have an augmented survivor bias given the COVID epidemic. Our institution normally has roughly 100 students per semester; for the Fall 2021 semester we had only 51 enrolled. We assumed these would have been those students most open to diversity. Finally, we assumed that the particular chronology of this semester would make it less likely that we would find significant changes in the aggregate measures; there seems to be general agreement in the field that longer experiences are more likely to lead to significant (and lasting) outcomes. Ward and Kennedy, talking generally about "sojourner adjustment" (i.e., getting over culture shock and integrating into a host culture on some level), measured the period in months, the flattening of the adjustment curve coming after six months (Ward & Kennedy, 1999). Some research confirms what international education professionals have long suspected: longer program lengths lead to more significant growth. One longitudinal study of study abroad over the last several decades suggested that the most growth came from the increasingly rare full-year students, while another showed that 13-18 weeks abroad was the ideal program length (Dwyer, 2004; Vande Berg et al., 2009). While we normally have a fourteen-week semester, we had compressed the term into just eleven weeks (ten weeks of instruction and one special events week) to obviate the need for a student visa. While we hoped for a change in at least some of the measures on the M-GUDS-S scale, we anticipated that this semester would be a pilot, to be redone during a "normal" semester. That

said, our research provides evidence that shorter programs (11 weeks, compared to Dwyer's "ideal" of 13-18) can have a significant impact on participants: this may be important in a post-pandemic world in which even semester programs may grow shorter.

The actual results were both pleasantly surprising and wide-ranging. We list the 15 questions we asked students in the figure below.

FIGURE 1

FALL 2021 SEMESTER COMPARISON (BEGINNING AND END) – REGULAR AND REVERSE SCORED

	Beginning		End	
	Mean	Std. Deviation	Mean	Std. Deviation
I know a lot of people from different countries	3,68	1,650	4,58	1,381
I would like to join an organization that emphasizes getting to know people from different cultures	5,15	1,026	5,40	1,056
Persons with disabilities can teach me things I could not learn elsewhere	5,36	0,922	5,56	0,873
I can best understand someone after I get to know how he/she is both similar to and different from me	4,89	1,068	5,23	1,134
I often listen to music of other cultures	3,74	1,595	4,33	1,492
Knowing how a person differs from me greatly enhances our friendship	4,57	1,201	5,13	1,178
I am interested in learning about the many cultures that have existed in the world	5,40	1,132	5,65	0,838
The real value of a college education lies in being introduced to different values	5,02	1,232	5,60	0,869
I attend events where I might get to know people from different cultural or national backgrounds	4,75	1,125	5,08	1,108
Knowing about the different experiences of other people helps me understand my own problems better	5,02	1,028	5,52	0,945
Grouped Means Regular Scored	4,76	n/a	5,21	n/a
Getting to know someone of another culture is generally an uncomfortable experience for me	2,23	1,250	2,29	1,543
I am not at ease with people of my nationality/culture	2,08	1,207	2,06	1,465
It's really hard for me to feel close to a person from another culture	1,94	1,064	1,96	1,352
It is very important that a friend agrees with me on most issues	2,68	1,173	3,19	1,439
I often feel irritated by persons of a different culture	1,57	0,772	1,87	1,453
Grouped Means Reverse Scored	2,10	n/a	2,27	n/a

The questions are divided into two categories: questions whose score was regular and those with reverse scoring. In other words, for the first ten questions, the range of responses went from strongly disagree (a response we converted into a numeric score of 1) to strongly agree (which was converted to a 6). A higher number corresponded to a higher UDO: for example, the responses students who responded "strongly agree" to the statement "I know a lot of people from other countries" were scored a six. The more agreement, the higher the numeric score. Five of the fifteen questions were however "reverse scored"—in other words, the scores students who strongly disagreed would result in a score of six for that

question. An example is "I often feel irritated by persons of a different culture." For someone to have a high M-GUDS-S score, they needed to respond strongly agree to the ten regularly scored questions and strongly disagree to the other five. Having this mix of regular and reverse scoring avoids what is called response bias, i.e., the tendency of people doing surveys to respond with a column of identical answers.

We examined both individual responses for their change over time as well as the average of the two categories or regular and reverse-scored responses. We hoped for significant increases in the regular-scored questions and a decrease in the average for the reverse-scored questions. For both semesters, many of the questions in the first group increased significantly, as did the overall average of that category. In Fall 2021, the average of regular-scored questions went from an initial score of 4.76 out of six to an end-of-semester value of 5.21, whereas in Spring 2022 the average rose from 4.74 to 4.94 (Figure 2 on the next page). Given both the brevity of the semester (eleven weeks), the epidemiological limits on meeting and spending time with others, and the possibility of a "ceiling effect" (i.e., that this populations' M-GUDS-S's scores would already be quite high), we were both surprised and very gratified to see this increase. Both increases in average score were significant, 9.4% in the Fall and 4.2% Spring. It's worth noting that we have had statistically significant aggregate increases in means every semester since starting this research in Fall 2021.

The data from Spring 2022 showed the same overall effect of significantly raising the average value of the regular-scored questions, with some differences in the individual questions. For instance, we found the same significant increases for the questions about getting to know people and music from other countries (increases of 3.56 to 4.22, $p < .05$, and 3.68 to 4.14, $p < .05$, respectively). However, for questions about how knowing a person's differences enhances friendship and being introduced to differences as the true value of a college education, the increases were much smaller than Fall 2021 and not statistically significant. One other question, about whether people with disabilities being able to teach us things we could not learn elsewhere, did not register a statistically significant increase. This followed the pattern from Fall 2021 but was disappointing given the presence of a course on Critical Disabilities in Spring 2022, one with a significant community engagement component involving our students working weekly with a local association teaching sports to neuroatypical people. A possible explanation was that only six students took that course, hence not enough perhaps to influence the average for that question.

FIGURE 2

SPRING 2022 SEMESTER COMPARISON (BEGINNING AND END) – REGULAR AND REVERSE SCORED

	Beginning		End	
	Mean	Std. Deviation	Mean	Std. Deviation
I know a lot of people from different countries	3,56	1,544	4,22	1,378
I would like to join an organization that emphasizes getting to know people from different cultures	4,87	0,859	5,05	1,094
Persons with disabilities can teach me things I could not learn elsewhere	5,32	0,842	5,33	0,839
I can best understand someone after I get to know how he/she is both similar to and different from me	4,82	1,039	5,06	1,017
I often listen to music of other cultures	3,68	1,397	4,14	1,456
Knowing how a person differs from me greatly enhances our friendship	4,74	0,964	4,90	1,020
I am interested in learning about the many cultures that have existed in the world	5,54	0,779	5,40	0,908
The real value of a college education lies in being introduced to different values	5,34	0,763	5,35	0,797
I attend events where I might get to know people from different cultural or national backgrounds	4,52	1,139	4,75	1,164
Knowing about the different experiences of other people helps me understand my own problems better	5,09	0,915	5,24	0,830
Grouped Means Regular Scored	4,74	n/a	4,94	n/a
Getting to know someone of another culture is generally an uncomfortable experience for me	2,22	1,218	2,15	1,305
I am not at ease with people of my nationality/culture	2,12	1,172	2,29	1,389
It's really hard for me to feel close to a person from another culture	2,02	1,145	1,98	1,204
It is very important that a friend agrees with me on most issues	3,45	1,167	3,22	1,321
I often feel irritated by persons of a different culture	1,71	1,078	1,60	1,051
Grouped Means Reverse Scored	2,36	n/a	2,24	n/a

It might be argued that while students showed a significant increase in their M-GUDS-S scores (both the average scores and for specific questions), that at this stage of their growth, *any* semester would have shown this sort of increase. In other words, undergraduate years are by definition a period of enormous personal growth and it was not the students' time studying at The Umbra Institute but rather simply the passage of time at this age. It is worth remembering that other studies using the M-GUDS have shown that not all multi-month experiences, even those very intentionally planned, lead to an increase in students' average score. In a notable example, Longerbeam and Sedlacek (2006) gave the measure to two populations of college students: one group was participating in a year-long civic living-learning program while the other group was a control made of undergraduates not taking part in the program. Despite the fact that the living-learning community students participated in series of events that were designed to increase their civic awareness and appreciation for diversity, there were no statistically significant differences in scores between the two populations (Longerbeam & Sedlacek, 2006). Another study in which an important period of

undergraduate life and learning did not lead to significant increases on M-GUDS scores was in the 2006-2007 Wabash National Study of Liberal Arts Education. First-year students—whom, it would be easy to imagine intuitively would be the most likely to have significant change in their appreciation of universalism and diversity—actually had a significant *decline* (at the $p = 0.05$ level) both in aggregate and across all three of the M-GUDS sub-measures (Blaich & Wise, 2007).

Another possible counterfactual might be that any heterogeneous environment—not necessarily study abroad at The Umbra Institute—might have led to a significant change in scores. Strauss and Connerley (2003) found, however, that there was no significant statistical association between living in a heterogeneous environment and having a more positive view on universality and diversity (the index used here was also the M-GUDS-S). The authors of that study concluded that while "contact may cause people to continue to seek diverse experiences with others, it will not ensure that people value or feel more connected to diverse others (p. 170). This is a further indication that it is the actual experience of studying at The Umbra Institute, rather than simply being in the city of Perugia, that is the driving force in the change we found in students.

4.2. Qualitative Results: Inside the Black Box of Students' Change During the Semester

The statistically significant change in the students' M-GUDS-S scores were a confirmation that study abroad at our program actually does change students, but it was important for us to understand what parts of the students' experience have been the drivers of that change. We evaluated a random (and what we believe is representative) sample of eighteen student journals. In them we found interesting insights on how the students lived their period abroad and the evolution in their perception of the intercultural differences, the change in their thoughts related to the others, and in their self-confidence. The journals were structured to accompany the students throughout their study abroad; despite being part of the Italian course, journal questions were in English and students —while free to respond in Italian—generally responded in English. During the first two weeks, students took their first steps with the Italian language and express their enthusiasm for Perugia, their new city.

Interesting results emerged from the analysis of the reflections students made during their first weeks of study. For the question on "my first impressions of Perugia", the comments are all similar: they express enthusiasm and excitement for the first days in a new city and the new people they are meeting. They reported exploring, wandering around, and discovering new corners of the city, restaurants, fashionable boutiques, gelato shops, and unique stores. They had big expectations for their study abroad, wanting to savor every minute, to learn,

grow, and discover their sense of direction. This question in the course journal maps on, to a certain extent, the tenth question in our modified M-GUDS-S scale: "I am interested in learning about the many cultures that have existed in the world."

There was a journal prompt dedicated to the new relationships that students were developing during study abroad. It is clear that in the majority of cases they were speaking of the friendships made with the other US students studying in the same program. They were really aware of how special the relationships they are creating are. Many said that there are common traits among them as they all chose to live a similar experience, there is a "common ground". All are welcoming and inclusive, outgoing and adventurous, and ready to meet new people and do something new. "We are all in this together, navigating this new world together, we are on the same boat here," someone said. Only one student explicitly mentioned friendships with students outside their program, "You allow yourself to be more open and open up to each other I have made a German guyfriend, some[one] from Spain who is on the Erasmus program, and some Italian girlfriends." Again here the comments suggest that students are reflecting on the themes that M-GUDS-S questions four and eight probe, "Getting to know someone of another culture is generally an uncomfortable experience for me," and "Knowing how a person differs from me greatly enhances our friendship."

The linguistic and cultural journal prompts in the final couple of weeks were more focused on what the students felt they had learned, as the end of the program was approaching. The responses were both personal and intimate: students reported knowing themselves better, both their positive qualities as well as their "bad" tendencies and habits. Many felt more independent and patient, were less afraid in general, and had learned to overcome bad times by talking to other and being true to themselves. They felt they had grown and learned how to be more social, more open to new experiences, and more comfortable in difficult situations and in dealing with different personalities. More than one mentioned that they have a broader perspective of the world and another frequent response was the discovery that there are other ways of living. One student wrote: "The people I have met opened my eyes to other people's issues," which reflects thought about the M-GUDS-S question 14, "Knowing about the different experiences of other people helps me understand my own problems better."

The analysis of the journals suggests a progressive transformation of the students throughout their study abroad experience. After the initial enthusiasm for the city's beauty, they started noticing differences that forced them to question their own identity. They started asking themself questions not so much about the host country but their own culture, habits, and traditions. This constant comparison process helped students understand that differences should not be

taken as a challenge to their own culture. Knowing about the different experiences of other people helped them understand their problems better. Through differences, they know themselves better, and they start loving themself and their origins even more. If the fundamental concept that the M-GUDS-S scale assesses is an openness to both diversity and universality, it is clear that the students' experience at The Umbra Institute and as a "temporary citizen" of Perugia helped them. As one student wrote, "Studying abroad…I have truly realized how different I am to others. I have met so many people from all over the world and it is interesting to see differences and/or similarities."

These qualitative results also made us more confident in the quantitative data from the M-GUDS-S scale. As Bowman notes, students are more accurate at self-assessment when "the outcome is salient and important to them" (Bowman, 2010, p. 67). The fact that so many students reported in the open-ended questions on the initial survey that some of their main goals for their time in Perugia were personal growth and meeting new people suggest that their self-report data is likely to be more accurate relative to the general student population; we think that we have avoided the main pitfalls in the use of self-report data (Kuh et al., 2001; Pascarella, 2001) while at the same time using qualitative sources to confirm and explain the quantitative results.

5. Concluding Discussion: Limitations, Implications, and Future Research

We undertook this research to be sure we were using correct (and sufficient) inputs paired with appropriate, impactful activities to result in better programmatic outcomes for our students. This pilot research study had certain limitations which however suggest some avenues for future research. One was the timing of the pretest: we wanted to conduct it as early as possible in the student experience as well as give it in a moment where we could be assured of standardized instruction and complete compliance on the part of the students. There are few moments in the academic calendar when all the students are in one place, other than the orientation small-group meetings run by the staff and the final community-engagement presentations. By using the staff as experimenters, we could assure a standardized set of instructions and a framing that emphasized the anonymity of the test.

The downside of this approach was that both moments (orientation meetings and final presentations) have a limited amount of time. We deemed it inopportune to give the students a lengthy survey especially because there were four open-ended questions at the end and we wanted the students to answer them fully, rather than with the telegraphic responses that often result for questions at

the end of a long survey. We therefore opted for the shorter M-GUDS-S scale rather than the full 45-question inventory. In addition, we used a modified version of the scale, which means that we need to re-run reliability and validity statistics on the revised measure. In future iterations of this research, we could also use the longer form of the M-GUDS scale; to see if there is any difference from these two semesters in which we used the short version. One other opportunity we have in the future would be to more intentionally calibrate the questions in the Italian reflection journal to the questions on the M-GUDS-S scale. The "progress" we believe is clear in the students' responses might have something to do with the students' transformation on the program, though this could also be due to the fact that the prompts in the journals were, each week, more focused on eliciting feedback about the students' experiences. The Italian reflection journal questions were not drafted with an eye to what the M-GUDS-S measures, and doing that in the future might help students to reflect on these significant aspects of universality and diversity. This all said, given the anonymity of the survey, there is no way to connect journal responses to specific students and their scores.

It seems reasonable to assume that the effects we found were due to the students' experience living and studying in Perugia. The lack of a control group, however, makes this nothing more than a reasonable hypothesis. We have therefore pondered trying to assemble a control group of students who resembled our students demographically but who had chosen not to study abroad. This could be to incentivize in some way each of our students to get a "buddy" who would take both the pre-test and the post-test while in the US. Alternatively, we could attempt to get students who had applied but not decided to study with us to fill out a survey at the beginning and end of a semester when they were back in the US and our students were here in Perugia, or perhaps even work with our institutional partners in the US to survey a control group. Another thing that would allow us to assess the long-term effects would be to attempt to have a second post-test several months after the students had returned to the US. The methodological advantages are obvious but the logistical-pragmatic obstacles are obvious to anyone who has attempted to conduct a longitudinal study: it is difficult to get compliance on re-tests after the subjects have, in a word, dispersed.

In the process of reviewing the literature after we had collected the data for the present study, we also came upon other variables for which we could in the future collect data. Basing his findings on the Wabash National Study, Bowman showed that while students are generally unable to estimate their learning and development accurately, first-generation college students are marginally better at this task (Bowman, 2010). Fuertes and Gelso (2000) also found gender differences, with women scoring significantly higher on the M-GUDS-S scale than men; there seemed to be variation both over time as well as within

populations. Somewhat surprisingly for research conducted in the year 2000: Fuertes and Gelso used the term "gender" and not "sex"; the advantages of the former over the latter are clear. Singley and Sedlacek (2004) found that American high school students in the top quartile of their graduating high school class had significantly higher ($p < .05$) average scores on the M-GUDS-S scale than their peers in the second quartile of the class. It would therefore be illuminating to assess whether or not academic performance correlates with UDO scores in our students. Another correlate that could be assessed in our future research is affect, specifically whether the sometimes negative feelings of frustration (especially linguistic barriers) and other hardships (the inevitable travail of weekend travel, adjusting to a new culture) ultimately have a positive effect on the students' M-GUDS-S scores (see Savicki, 2013). There are scattered mentions in the journals of students being frustrated, but we could be more intentional about asking about emotions and their effects.

All this said, immersion itself does not mean, *ipso facto*, greater cultural awareness. There are many factors that affect whether immersion in a culture leads to greater cross-cultural understanding and acceptance of difference (Wilkinson, 1998). Laubscher suggested over thirty years ago that true understanding of the host culture required "decisive intervention," either by a host-country informant or someone familiar with both countries (Laubscher, 1994, p. 107). It might be interesting to introduce a more intentional weekly forum in the students' Italian courses, or perhaps to assess whether those students who participate in an optional "conversation partner" program with Italian university students have different scores. As part of the Institute's agreements with the two local Italian universities, we have between ten to twenty Italian university students taking our courses alongside our American students. We would like to see whether the effect on UDO scores would be the same for these students, in other words if Italians "studying abroad" while at home would also experience a rise in scores.

Ultimately, we undertook this research with an eye towards iterating our program's various components to increase our students' immersion in and integration with the local culture. The results suggested that our program does indeed have a (statistically) significant impact on students' UDO score and that our efforts at promoting integration are an important part of that. In the past semester's survey (i.e., Spring 2023), we introduced other questions, aimed at disaggregating the observed effect, i.e., to determine which of the various parts of the student experience had the greatest impact on student growth. This updated survey had questions about the number of classes students had with a significant community engagement component, whether they had chosen housing with Italian and international students, which of our several curricular tracks (which

include facilitated direct enrollment in the two local Italian universities) they had chosen, the number of independent trips they took, etc. The results of this updated survey will be reported in a future article. We are in the midst of a curricular review to increase what Umbach and Kuh (2006) called "diversity press," which they wrote "represents the institution's commitment to and emphasis on diversity as manifested by the proportional presence of students from different backgrounds attending the institution (structural diversity), the extent to which students perceive that diversity is valued and important, and the degree to which diversity is featured in the curriculum" (p. 177).

We also realize that the collection of better demographic data and other variables would help us determine which students might need even more interaction. For example, we might see that students from small regional colleges who are first-generation students might score lower at the beginning of the semester. Given both the disproportionate effects of the COVID-19 pandemic on women (Nair 2022) and the overwhelming preponderance of women in our program, we might find that we need to do more to support certain activities or interactions. This might also be true for our students of color: already underrepresented in American study abroad programs, African-American participation in study abroad has dipped significantly during COVID and in the post-pandemic era (JBHE, 2021). In the short term, the results convinced us to "thicken" the community engagement components in our existing CE-related courses as well as to revise our Italian language program to become an even more intentional vehicle of cultural immersion. In the summer of 2021, our three Italian language professors worked with three recent graduates of a masters program in teaching Italian as an L2 to write an Italian 101 textbook specifically for our students. The results of this study certainly informed that writing and we will continue to iterate the textbook each year as we learn more about our students' process of growing while they live abroad and study with us. We will continue with the pre- and post-tests each semester and also explore the literature on successful living-learning programs, as given our program's structure, this is what The Umbra Institute most resembles (Inkelas et al., 2006).

In sum, this pilot research provides a foundation for further investigation into the growth of students studying abroad at The Umbra Institute and their openness to diversity. The progression within the students' journals and increase in their scores from the shortened M-GUDS-S survey demonstrates positive changes in open-mindedness, which contrary to personality, can be measured and examined as a result of its amendability (Yakunina et al., 2012). The push and pull of living in multicultural realities and a globalized world continues to prompt the need for dependable assessments to express changes in open-mindedness regarding diversity for students studying abroad. Ultimately, this research

demonstrates, through the use of a measurable and reliable scale, the achievement of growth in cultural competency and engagement with diversity that comes with an educational experience abroad.

Acknowledgments

The authors would like to thank the editors of *Frontiers* and the three anonymous reviewers, our colleague Elgin K. Eckert, and the students who participated in the surveys and focus groups, as well as ISI Abroad and Arcadia University's College of Global Education for supporting this research.

References

Ahlburg, D. A. (2020). COVID-19 and UK Universities. *The Political Quarterly*, *91*(3), 649–654. https://doi.org/10.1111/1467-923X.12867

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*(4), 457–469. https://doi.org/10.1016/j.ijintrel.2005.10.004

BEVI. (2023). *About the BEVI*. https://thebevi.com/about/

Bilecen, B. (2020). Commentary: COVID-19 Pandemic and Higher Education: International Mobility and Students' Social Protection. *International Migration*, *58*(4), 263–266. https://doi.org/10.1111/imig.12749

Blaich, C., & Wise, K. (2007). *Overview of Findings from the First Year of the Wabash National Study of Liberal Arts Education*. Gustavus Adolphus College.

Bowman, N. A. (2010). Assessing learning and development among diverse college students. *New Directions for Institutional Research*, *2010*(145), 53–71. https://doi.org/10.1002/ir.322

Braskamp, L. A., Braskamp, D. C., & Merrill, K. (2009). Assessing Progress in Global Learning and Development of Students with Education Abroad Experiences. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *18*(1), Article 1. https://doi.org/10.36366/frontiers.v18i1.256

Cozien, C. (2015). *Racism under the linguistic microscope: A preliminary investigation into what the critical analysis of South African news media discourse can reveal about racial stereotyping, and the relationship between culture and language* [Honours in Linguistics, University of Cape Town]. http://rgdoi.net/10.13140/RG.2.1.4580.5528

Deardorff, D. K. (2015). *Demystifying Outcomes Assessment for International Educators: A Practical Approach*. Stylus Publishing.

Deardorff, D. K., & van Gaalen, A. (2021). Assessing Internationalization Outcomes. In B. Leask, D. K. Deardorff, H. de Wit, & H. Charles (Eds.), *The Handbook of International Higher Education* (Second). Stylus Publishing, LLC.

Dwyer, M. M. (2004). More Is Better: The Impact of Study Abroad Program Duration. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*(1), Article 1. https://doi.org/10.36366/frontiers.v10i1.139

Engberg, M. E. (2013). The Influence of Study Away Experiences on Global Perspective-Taking. *Journal of College Student Development*, *54*(5), 466–480.

Engle, L. (2013). The Rewards of Qualitative Assessment Appropriate to Study Abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *22*(1), Article 1. https://doi.org/10.36366/frontiers.v22i1.321

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. https://doi.org/10.36366/frontiers.v9i1.113

Fitzgerald, E. A., Marzalik, P. R., & Kue, J. (2018). Assessing Intercultural Development Pre- and Post Education Abroad. *The Journal of Nursing Education*, *57*(12), 747–750. https://doi.org/10.3928/01484834-20181119-08

Fuertes, J. N., & Gelso, C. J. (2000). Hispanic counselors' race and accent and Euro Americans' universal-diverse orientation: A study of initial perceptions. *Cultural Diversity and Ethnic Minority Psychology*, *6*(2), 211–219. https://doi.org/10.1037/1099-9809.6.2.211

Fuertes, J. N., Miville, M. L., Mohr, J. J., Sedlacek, W. E., & Gretchen, D. (2000). Factor structure and short form of the Miville-Guzman Universality-diversity scale. *Measurement and Evaluation in Counseling and Development*, *33*(3), 157.

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.

Grant, J., Acheson, K., & Karcher, E. (2021). Using the BEVI to Assess Individual Experience to Enhance International Programming. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *33*(1), Article 1. https://doi.org/10.36366/frontiers.v33i1.491

Hammer, M. R. (2012). The Intercultural Development Inventory: A New Frontier in Assessment and Development of Intercultural Competence. In M. Vande Berg, R. M. Paige, & K. H. Lou (Eds.), *Student Learning Abroad: What Our Students Are Learning, What They're Not, and What We Can Do About It* (pp. 137–161). Stylus Publishing.

Inkelas, K. K., Vogt, K. E., Longerbeam, S. D., Owen, J., & Johnson, D. (2006). Measuring Outcomes of Living-Learning Programs: Examining College Environments and Student Learning and Development. *The Journal of General Education*, *55*(1), 40–76.

Iowa State University of Science and Technology. (2023). *About the GPI*. Global Perspective Inventory. https://www.gpi.hs.iastate.edu/

JBHE. (2021, November 2022). The Black Percentage of All U.S. Students Who Studied Abroad Took a Big Hit From the Pandemic. *The Journal of Blacks in Higher Education*. https://www.jbhe.com/2021/11/the-black-percentage-of-all-u-s-students-who-studied-abroad-took-a-big-hit-from-the-pandemic/

Klein, G. B., & Ravenda, A. F. (2016). Le parole che discriminano. In *Addio alla razza: Una parola pericolosa che per la scienza non ha senso* (Vol. 27/28, pp. 39–49).

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., & Kennedy, J. (2001). *NSSE technical and norms report*. https://scholarworks.iu.edu/dspace/handle/2022/24158

Laubscher, M. R. (1994). *Encounters with difference: Student perceptions of the role of out-of-class experiences in education abroad*. Greenwood Press.

Longerbeam, S. D., & Sedlacek, W. E. (2006). Attitudes Toward Diversity and Living-Learning Outcomes Among First- and Second-Year College Students. *NASPA Journal*, *43*(1), 40–55. https://doi.org/10.2202/1949-6605.1570

Miville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Holloway, P., & Fuertes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman Universality-Diversity Scale. *Journal of Counseling Psychology*, *46*(3), 291–307. https://doi.org/10.1037/0022-0167.46.3.291

Nevado Llopis, A., & Sierra Huedo, M. L. (2022). Promoting the Development of Intercultural Competence in Higher Education Through Intercultural Learning Interventions. *Revista Educare*, *26*(2), 1–21. https://doi.org/10.15359/ree.26-2.28

Norris, E. M., & Dwyer, M. M. (2005). Testing Assumptions: The Impact of Two Study Abroad Program Models. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *11*(1), Article 1. https://doi.org/10.36366/frontiers.v11i1.154

Pascarella, E. T. (2001). Using student self-reported gains to estimate college impact: A cautionary tale. *Journal of College Student Development*, *42*(5), 488.

Rogers, A. (2020). *Internationalizing the Campus at Home: Campus Globalization in the Context of COVID-19*. iieNetwork. https://iie.widen.net/s/dgz5cmjb6z/internationalizing-the-campus-at-home-1

Salisbury, M., An, B. P., & Pascarella, E. T. (2013). The Effect of Study Abroad on Intercultural Competence Among Undergraduate College Students. *Journal of Student Affairs Research and Practice*, *50*(1), 1–20. https://doi.org/10.1515/jsarp-2013-0001

Savicki, V. (2013). The Effects of Affect on Study Abroad Students. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *22*(1), Article 1. https://doi.org/10.36366/frontiers.v22i1.323

Savicki, V., & Brewer, E. (2015). Introduction: Issues in Assessing Study Abroad. In V. Savicki & E. Brewer (Eds.), *Assessing Study Abroad: Theory, Tools, and Practice* (pp. 1–12). Stylus Publishing.

Schenker, T. (2019). Fostering Global Competence through Short-Term Study Abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *31*(2), Article 2. https://doi.org/10.36366/frontiers.v31i2.459

Seo, M., & Konishi, T. (2023). Internal Changes in Students who Participated in an Online Short-Term Study Abroad Program: An Analysis Using BEVI and Interview. *Journal of International Student Advisors and Educators*, *25*, 63–76. https://doi.org/10.57511/coisan.25.0_63

Shealy, C. N. (Ed.). (2016). *Making sense of beliefs and values: Theory, research, and practice*. Springer Publishing Company.

Singley, D. B., & Sedlacek, W. E. (2004). Universal-Diverse Orientation and Precollege Academic Achievement. *Journal of College Student Development*, *45*(1), 84–89. https://doi.org/10.1353/csd.2004.0015

Stebleton, M. J., Soria, K. M., & Cherney, B. T. (2013). The High Impact of Education Abroad: College Students' Engagement in International Experiences and the Development of Intercultural Competencies. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *22*(1), Article 1. https://doi.org/10.36366/frontiers.v22i1.316

Stevens, D. D., & Cooper, J. E. (2009). *Journal Keeping: How to Use Reflective Writing for Learning, Teaching, Professional Insight and Positive Change*. Stylus Publishing.

Strauss, J. P., & Connerley, M. L. (2003). Demographics, Personality, Contact, and Universal-Diverse Orientation: An Exploratory Examination. *Human Resource Management*, *42*(2), 159.

Sutton, R. C., & Rubin, D. L. (2004). The GLOSSARI Project: Initial Findings from a System-Wide Research Initiative on Study Abroad Learning Outcomes. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*(1), Article 1. https://doi.org/10.36366/frontiers.v10i1.133

Umbach, P. D., & Kuh, G. D. (2006). Student Experiences with Diversity at Liberal Arts Colleges: Another Claim for Distinctiveness. *The Journal of Higher Education*, *77*(1), 169–192. https://doi.org/10.1353/jhe.2006.0008

Vande Berg, M., Connor-Linton, J., & Paige, R. M. (2009). The Georgetown Consortium Project: Interventions for Student Learning Abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *18*(1), Article 1. https://doi.org/10.36366/frontiers.v18i1.251

Vande Berg, M., Paige, R. M., & Lou, K. H. (Eds.). (2012). *Student Learning Abroad: What Our Students Are Learning, What They're Not, and What We Can Do About It*. Stylus Publishing.

Wandschneider, E., Pysarchik, D. T., Sternberger, L. G., Ma, W., Acheson, K., Baltensperger, B., Good, R. T., Brubaker, B., Baldwin, T., Nishitani, H., Wang, F., Reisweber, J., & Hart, V. (2015). The Forum BEVI Project: Applications and Implications for International, Multicultural, and Transformative Learning. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *25*(1), Article 1. https://doi.org/10.36366/frontiers.v25i1.350

Ward, C., & Kennedy, A. (1999). The measurement of sociocultural adaptation. *International Journal of Intercultural Relations*, *23*(4), 659–677. https://doi.org/10.1016/S0147-1767(99)00014-0

Whatley, M., Landon, A. C., Tarrant, M. A., & Rubin, D. (2021). Program Design and the Development of Students' Global Perspectives in Faculty-Led Short-Term Study Abroad. *Journal of Studies in International Education*, *25*(3), 301–318. https://doi.org/10.1177/1028315320906156

Wilkinson, S. (1998). On the Nature of Immersion During Study Abroad: Some Participant Perspectives. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *4*(1), Article 1. https://doi.org/10.36366/frontiers.v4i1.65

Yakunina, E. S., Weigold, I. K., Weigold, A., Hercegovac, S., & Elsayed, N. (2012). The multicultural personality: Does it predict international students' openness to diversity and adjustment? *International Journal of Intercultural Relations*, *36*(4), 533–540. https://doi.org/10.1016/j.ijintrel.2011.12.008

Author Biography

Zachary Nowak has worked in study abroad since 2005, in roles as varied as Student Services Assistant, US University Relations, and Director of The Umbra Institute in central Italy. He received his PhD in 2018 and is interested in cultural geography and pedagogy; he has recently been doing research on the assessment of study abroad.

Jacopo Bernardini is one of the Umbra Program Coordinators. He holds a PhD in Social and Political Theory and Research and has worked for six years as a Postdoctoral Researcher of Sociology and Methodology of Social Research at the Faculty of Political Science of the University of Perugia. He has published several academic articles and books on Age and Life Cycle, Transition to Adulthood, Generational Conflicts, and the Information Age.

Paola Cascinelli is the director of the Italian branch of Arcadia University and teaches organizational behavior in multicultural contexts. She has been involved in training for personal, social, and organizational development. She has a passion for the Mediterranean and its potential, carefully looking at the interaction between the countries that overlook the white sea. Since the beginning of her career, Paola has undertaken independent research, connecting the theoretical debate with her professional life.

Amanda Fontana is an Admissions Specialist in ISI Abroad's Boston Office. She is pursuing a Global Master's of Arts in International Education at SIT Graduate Institute. She was an intern at The Umbra Institute working with the Community Engagement and Student Services departments and is now an Enrollment Management and Institutional Relations Assistant for ISI Abroad. She holds a BA in African American and African studies from The Ohio State University. She is passionate about the future of Americans studying abroad and becoming global citizens.

Appendix A. The Umbra Institute's version of the Miville-Guzman Universality-Diversity Scale – Short Form

The reverse-scored questions are in italics.

Response options for each item: 1 = strongly disagree, 2 = disagree, 3 = disagree a little bit, 4 = agree a little bit, 5 = agree, 6 = strongly agree

1. I know a lot of people from different countries.
2. I would like to join an organization that emphasizes getting to know people from different cultures.
3. Persons with disabilities can teach me things I could not learn elsewhere.
4. *Getting to know someone of another culture is generally an uncomfortable experience for me.*
5. I can best understand someone after I get to know how they are both similar to and different from me.
6. *I am not at ease with people of my nationality/culture.*
7. I often listen to music of other cultures.
8. Knowing how a person differs from me greatly enhances our friendship.
9. *It's really hard for me to feel close to a person from another culture.*
10. I am interested in learning about the many cultures that have existed in the world.
11. The real value of a college education lies in being introduced to different values.
12. *It is very important that a friend agrees with me on most issues.*
13. I attend events where I might get to know people from different cultural or national backgrounds.
14. Knowing about the different experiences of other people helps me understand my own problems better.
15. *I often feel irritated by persons of a different culture.*

Appendix B. Additional Methodological Notes

Given the potential utility to other study abroad programs of a valid scale of assessment, it was important to us to include some more specific notes on our methodology for those interested in replicating it. There were different moments of data collection and kinds of data, as we gathered both quantitative and qualitative information. On their first full day in Italy as part of their orientation program, students used a QR code to access pre-test (T1), a Google form. This form had both the fifteen questions of the short version of the Miville-Guzman Universality-Diversity Scale (M-GUDS-S), as well as a number of open-ended questions about the equivalencies for their Italian course and their expectations for their experience in Perugia and at The Umbra Institute.

As noted above, we processed the quantitative data using IBM's SPSS Statistics software. We first created and cleaned the data matrix, then transformed the string variables from the matrix extracted from the Google form into numeric variables. We performed an analysis of all variables, collating descriptive statistics and carrying out crosstabulation between scales and most significant demographic and non-demographic variables (gender, age group, Italian language learning priority, type of home institution, Federal Pell Grant recipient, hours spent on co- or extracurricular activities). We then compared the most significant results from the two surveys (beginning and end of semester) for both semesters in question, followed by the calculation and analysis of standard deviation and kurtosis/skewness distribution for all variables.