Selling America:

U.S. Public Diplomacy Programs in the Middle East and South Asia in a Post 9/11 Environment

Elana Berger

Public Policy Honors Thesis

Advisors: Professors Fritz Mayer and Ken Rogerson

Sanford School of Public Policy, Duke University

December 4, 2009
Table of Contents

Acknowledgements 3
Abbreviations Glossary 4

I. What is public diplomacy and why is it important? 5

II. Public diplomacy: An essential component of successful international relations 9

III. Methodology 24

IV. Empirical analysis: Applying these criteria to evaluations of public diplomacy programs 39

V. Conclusion 80

VI. Works Cited 88
Acknowledgements

I want to thank everyone who has supported and guided me throughout this thesis-writing process. First, I would like to thank both my advisor, Professor Fritz Mayer and my Honors Seminar Professor, Ken Rogerson, for advising me, answering my questions every step of the way and for critically reading many versions of my research. I really appreciate all of the time and effort you put into this project to make it a success. I also want to thank my parents and my sister for supporting me throughout my college career at Duke and throughout the time I spent working on this research project. I especially want to thank my dad for offering new, unique perspectives on my research topic and my mom for reading numerous drafts of my thesis, always providing invaluable comments throughout the entire process. Also, thanks to my peers, my fellow Public Policy Honors Seminar students, who provided their suggestions and support throughout the past year. Finally, thanks to all of my friends at Duke who have always been there for me when I needed them throughout these past four years.
Abbreviations Glossary

**ECA:** Bureau of Educational and Cultural Affairs, U.S. Department of State

**IIP:** Bureau of International Information Programs, U.S. Department of State

**R/PPR:** Office of Planning, Policy and Resources; U.S. Department of State

**PDEO:** Public Diplomacy Evaluation Office; combines the evaluations staff of ECA, IIP, and R/PPR; U.S. Department of State

**ECA/P:** Office of Policy and Evaluation, Bureau of Educational and Cultural Affairs; U.S. Department of State

**EMU:** Evaluation Measurement Unit; Office of Planning, Policy and Resources; U.S. Department of State

**BBG:** Broadcasting Board of Governors; an independent (non-governmental) entity in charge of international nonmilitary broadcasting

**GAO:** U.S. Government Accountability Office; supports the U.S. Congress and helps improve the performance and accountability of the federal government

**Access/Access Program:** The English Access Microscholarship Program

**BRIDGE/BRIDGE Program:** Building Respect through Internet Dialogue and Global Education Program

**SCP:** School Connectivity Program

**IV/IVP:** International Visitor Program: Special Initiative Projects in Response to the Events of September 11, 2001

**MBN:** Middle East Broadcasting Network (of which Alhurra TV is a part)
I. What is public diplomacy and why is it important?

Since the terrorist attacks on the World Trade Center in New York City on September 11, 2001, foreign policy experts and U.S. government officials have emphasized the importance of public diplomacy in combating terrorism and ensuring national security. In the current climate of anti-American sentiment, “the war of ideas is more challenging than ever” and strong negative public opinion about the United States is pervasive (Pilon). According to the 2007 Pew Global Attitudes Project, while majorities in 25 of the 47 countries surveyed express positive views of the United States, since 2002, the image of the United States has declined in most parts of the world. Favorable ratings of America are lower in 26 of 33 countries for which trends are available. The U.S. image remains abysmal in most Muslim countries in the Middle East and Asia; in 2007, favorable views of the United States were at just 9% in Turkey and had declined to 15% in Pakistan (The Pew Global Attitudes Project: Global Unease With Major World Powers Rising Environmental Concern in 47-Nation Survey). While the election of Barack Obama and favorable views towards the new president have allowed for the U.S. image to rebound in some places, the effect in predominantly Muslim nations has been insignificant thus far. While Obama is generally much more popular abroad than was his predecessor, there is still a high level of distrust and animosity towards the United States in many Muslim nations (Wike).

Public diplomacy attempts to combat this anti-American sentiment, with a special focus on those Arab and Muslim populations where attitudes about America are particularly negative. Public diplomacy, “the promotion of America’s interests, culture and policies by informing and influencing foreign populations,” includes three categories of activities: international information programs, educational and cultural exchanges, and international nonmilitary broadcasting. The first two categories fall under the purview of the State Department, the third is the responsibility
of the Broadcasting Board of Governors (BBG), an organization independent from the federal government (Epstein and Mages). Many government officials and foreign policy experts have recently critiqued several programs, organized by these institutions, for not living up to their potential. It is certainly an ominous sign for American public diplomacy programs when Defense Secretary Robert Gates admits “that al-Qaeda is better at communicating its message on the internet than America” (Pilon 6).

Numerous foreign policy experts suggest changes to the methods, structures, and goals of public diplomacy programs, which will theoretically lead to better understanding of American policies and improve America’s image abroad. Nevertheless, few of the reports critiquing public diplomacy programs mention evaluations of specific international information programs, educational and cultural exchanges, or international non-military broadcasting efforts. This omission is largely because, until just a few years ago, there were few evaluation methods in place to assess the impact of public diplomacy, especially at the Bureau of International Information Programs (IIP), making it nearly impossible to judge whether these programs are working (ExpectMore.gov "Detailed Information on the International Information Programs Assessment"). The Bureau of Educational and Cultural Exchanges (ECA) has had a “culture of measurement” in place for at least a decade, but only with the creation of the Office of Policy, Planning and Resources for Public Diplomacy and Public Affairs (R/PPR) and the Public Diplomacy Evaluation Office (PDEO) in 2004, has IIP attempted to imitate these measurement techniques and apply them to their own programs ("U.S. Department of State Bureau of Educational and Cultural Affairs: Public Diplomacy Evaluation Office "; "Evaluation and Measurement Unit: Office of Policy, Planning and Resources for Public Diplomacy and Public
The PDEO dissolved in 2008 after Karen Hughes, former Under Secretary for Public Diplomacy and Public Affairs, left the State Department, but the new Evaluation and Measurement Unit (EMU) at R/PPR is essentially a replacement for PDEO, and is “responsible for conducting impact assessments and evaluations for all non-exchange public diplomacy programs” (Brewer). The EMU oversees four major areas: 1. the Public Diplomacy Impact Project, which uses surveys and focus groups to measure the impact of public diplomacy programs on foreign audiences; 2. program evaluations to analyze the long-term impact of public diplomacy programs; 3. The Mission Activity Tracker, an online performance measurement tool that records the scope, frequency and efficacy of U.S. public diplomacy activities; and 4. The Performance Based Budgeting Pilot, an initiative that evaluates whether U.S. programs are achieving their objectives with the current allotted funding for U.S. public diplomacy programs, determining if new strategies or additional funds are necessary to reach key public diplomacy objectives ("Evaluation and Measurement Unit: Office of Policy, Planning and Resources for Public Diplomacy and Public Affairs (R/PPR)"). The BBG has also recently improved and expanded their evaluation efforts by measuring audience size and credibility for their programs (ExpectMore.gov "Detailed Information on the Broadcasting in Arabic Assessment"). It is now feasible to determine how public diplomacy programs impact the foreign audiences who directly participate in these programs.

While there is some disagreement among foreign policy experts and government officials about how to best structure and improve public diplomacy programs, there is a general consensus that effective public diplomacy is essential to national security and the war on terrorism.
Nevertheless, although many critics cite public opinion polls as evidence for the inefficacy of public diplomacy programs, it is not actually possible to know whether or not these programs are accomplishing their goals without proper evaluation methods that tie measurable results directly to these programs. Therefore, determining whether or not there are effective evaluation techniques in place to measure the results of these programs is the necessary first step to assessing whether or not the programs are accomplishing their goals.

There have been many different programs, strategies and messages used since the September 11th attacks in 2001, in U.S. public diplomacy efforts in predominantly Muslim countries in the Middle East and South Asia. I analyzed evaluations of four specific public diplomacy programs in this region to address the following question: How are public diplomacy programs evaluated and do evaluations of these programs accurately reflect the quality of the programs? I chose these four programs because they fit into the scope of my research, as far as geographic region and time frame, and because complete evaluations of these programs were publicly-available. Since my research focuses on the quality of the evaluations of public diplomacy programs and not the efficacy of the programs themselves, I needed to choose programs for which complete, thorough evaluations were available. There were very few programs that met all of these criteria.

This study could have important policy implications, leading to recommendations about which evaluation methods and programs are functioning and which are not, and if new types of measurement techniques should be implemented. Additionally, with the change in administrations, it is possible that techniques that were ineffective under the Bush administration may be more effective under the new policy and political environment of the Obama administration.
II. Public diplomacy: An essential component of successful international relations

A) Anti-American sentiment measured in global public opinion polls

The 2007 Pew Global Attitudes Survey seems to provide evidence that effective public diplomacy programs may be more important than ever before, particularly in the Middle East and South Asia. The survey shows that negative views of the United States continue to prevail in most of the Muslim world. These results are particularly troubling because this high level of anti-Americanism is likely to have negative effects on U.S. economic interests around the world, to limit the ability of the government to pursue U.S. foreign policy, and to decrease the security of Americans around the world (Ford 6-7).

On the bright side, anger towards the United States in the Arab and Muslim world is less universal today than it was in 2003, immediately after the start of the war in Iraq. Nonetheless, positive opinions held by foreigners about the American people declined between 2002 and 2007, in 23 of the 33 countries where trends are available. In both Indonesia and Turkey, favorable views of Americans have declined 23 and 19 percentage points respectively between 2002 and 2007 (The Pew Global Attitudes Project: Global Unease With Major World Powers: Rising Environmental Concern in 47-Nation Survey 4). The Pew Research Center cites various reasons for America’s generally negative image abroad, including the belief that U.S. policies increase the gap between rich and poor nations, a backlash against American values and the spread of American ideas and customs, and a dislike of the American way of doing business.

Notably, despite their generally negative opinion of the United States, Middle Easterners actually tend to favor the way Americans do business. However, majorities in several predominantly Muslim countries disapprove of the spread of American popular culture to their countries (The
Additionally, in 30 out of 34 countries where trends are available, support for the U.S.-led war on terrorism declined since the 2002 poll. Even in several countries that have experienced terrorist attacks in recent years, such as Indonesia, Bangladesh, Spain, Jordan, Morocco, Pakistan and Turkey, majorities say they are opposed to America’s war on terrorism (The Pew Global Attitudes Project: Global Unease With Major World Powers Rising Environmental Concern in 47-Nation Survey 6).

Although support for Bin Laden and violent extremism has decreased in many predominantly Muslim countries in the past few years, there is still substantial support (between 25% and 60%) in countries such as Jordan, Pakistan, Lebanon, Morocco, and Indonesia for both Bin Laden and for the use of suicide bombings against the United States and its allies (The Pew Global Attitudes Project: Islamic Extremism: Common Concern for Muslim and Western Publics Support for Terror Wanes Among Muslim Publics). Although there are numerous factors that contribute to these negative opinions and perceptions, many scholars and government officials find it difficult to believe that U.S. public diplomacy programs are working if opinions about America and its war on terrorism are still so negative (Pilon; "Public Diplomacy in the Middle East and South Asia."); Epstein and Mages).

The Pew study also highlights that “familiarity breeds favorability.” In general, respondents are more likely to have favorable attitudes towards America if they have visited the United States or if they have family or friends in the United States. This may also partly explain why attitudes toward Americans are more positive than attitudes toward the United States as a nation (The Pew Global Attitudes Project: Global Unease With Major World Powers Rising Environmental Concern in 47-Nation Survey 18-19). When foreigners develop personal
relationships with Americans, they are more likely to make the distinction between the American people and the American government. This statement provides a strong case for the importance of educational and cultural exchanges and “cultural diplomacy,” the development of respect for and understanding of other cultures and their way of thinking (Schneider "Cultural Diplomacy: Hard to Define, but You'd Know It If You Saw It" 192).

B) Overview of the most common recommendations for public diplomacy programs

In the past 8 years, the public diplomacy efforts of both the State Department and the BBG have been examined and experts have provided an array of recommendations for ways to improve these programs. A review of 29 of the most credible reports written on public diplomacy between 2002 and 2005 reveal that the most common recommendations for public diplomacy programs include: increase exchanges and libraries, increase public diplomacy and language training, increase financial and human resources, define overall strategy, increase private sector involvement, and create a presidential directive or reorganize public diplomacy at the White House (Epstein and Mages 4). Other recommendations include: create a new agency, increase technology use, reorganize at the State Department, redefine the role of the Under Secretary for Public Diplomacy and Public Affairs, increase embassy involvement, and improve communication (Epstein and Mages 4). Additionally, many public diplomacy experts also urge the State Department and the BBG to develop a more culturally-based approach to public diplomacy and to increase oversight and evaluation of public diplomacy programs (Schneider "Cultural Diplomacy: Hard to Define, but You'd Know It If You Saw It"; Epstein and Mages 4; Dutta-Bergman; Schneider "Culture Communicates: US Diplomacy that Works")
a) Increase foreign language training

With regards to the need for better foreign language training, a 2006 Government Accountability Office (GAO) study indicates that 30 percent of the language designated-posts in the Muslim world are filled by officers without the requisite language skills (U.S. Public Diplomacy: State Department Efforts to Engage Muslim Audiences Lack Certain Communication Elements and Face Significant Challenges 2). Administrators at the major public diplomacy institutions also identify the lack of sufficient language training for public diplomacy employees, particularly in Arabic, as one of the major challenges still facing their efforts ("Public Diplomacy in the Middle East and South Asia."). If these Foreign Service Officers are not fluent in Arabic or other local languages, it is unclear how they can effectively communicate with people in the Muslim world.

b) Increase financial resources

Increasing funding for public diplomacy might also help to improve the efficacy of the programs. Public diplomacy efforts are underfunded given their importance to national security and fighting terrorist ideology. The amount of money devoted to public diplomacy programs pales in comparison to the military budget (Epstein and Mages; Peterson).

c) Define overall strategy and clarify the purpose of public diplomacy

The American strategy for public diplomacy efforts and the purpose of these efforts may need to be re-visited and clarified. For example, Peter Peterson of the Council on Foreign Relations explains that the perceived lack of empathy among Americans towards the pain, hardship, and tragic plight of people in the developing world is a major cause of negative attitudes towards the United States. However, expressing empathy, particularly if it appears to be contradictory to U.S. policies and values, will not by itself be enough (Peterson). Peterson
highlights that animosity is related to policy issues and that the United States cannot and should not always please others with its policy choices. Part of the challenge of effective public diplomacy is to better explain why the United States does what it does and then accept that others choose to differ. As Peterson explains, “some of the hostility can be offset but not eradicated. The United States should not leave an impression that all differences are resolvable or could be if it were just nicer or more empathetic. This is part of being a great power” (Peterson). Shawn Zeller makes a similar point in his article in the *Foreign Service Journal*; he explains the daunting task of “turn[ing] the tide of public opinion in the Muslim world” when U.S. policy has been fundamentally at odds with public opinion in many predominantly Muslim countries (20). For example, if there is solid public backing amongst Americans for American support of Israel, the U.S. government should not become less supportive of Israel simply because people in the Muslim world tend to disagree with this support. Nevertheless, the U.S. government must clearly articulate its values and policies and explain them to foreign audiences.

d) **Coordinate with allies and bridge the gap between the public and private sectors**

Peterson suggests that the United States should cooperate and coordinate with allies, like the U.K., whose sources (like the BBC) are often seen as more credible and objective than U.S. sources. Peterson also argues for the creation of an independent, not-for-profit “Corporation for Public Diplomacy,” to bridge the gap between public and private sector initiatives. This independent organization would be able to leverage the creativity and flexibility of the private-sector, receive private-sector grants, and benefit from the ability of private sector media to communicate the values and the merits of democracy more effectively than government officials. The establishment of such an institution would create a non-government messenger, and perhaps a more trusted messenger (Peterson).
e) Reorganize public diplomacy at the State Department

Others have also suggested changes to the organization of public diplomacy programs at the State Department. Cynthia Schneider, an expert in cultural diplomacy, emphasizes the need for a reliable, non-governmental messenger in public diplomacy programs. She argues that the current structure under the State Department compromises the independence of public diplomacy, which is essential to its credibility. According to Schneider, it is crucial to create distance between the cultural presence and the U.S. government for foreign audiences to embrace cultural diplomacy efforts. She stresses the need for more personal interactions in public diplomacy that involve people and not governments (Schneider "Cultural Diplomacy: Hard to Define, but You’d Know It If You Saw It").

f) Improve communication and dissemination of information about U.S. public diplomacy programs within the United States

Juliana Geran Pilon of the Heritage Foundation has a slightly different criticism of public diplomacy programs. She highlights the restrictions of the Smith-Mundt Act of 1948, which prohibits agencies from disseminating information that is meant for foreign audiences, within the United States. Upon request, this information can be made available to members of Congress, representatives of the press, or researchers, but is generally supposed to be unavailable to the general public. This act was originally designed to prevent the distribution of propaganda within the United States. On the other hand, the act also makes it difficult for the U.S. taxpayer to know what type of programming the State Department and BBG is providing in other countries. According to Pilon, the act is outdated and impractical in the current global information age. Because of the vast amount of information now available on the internet, the Smith-Mundt Act has already become ineffective.
Due to the Smith-Mundt Act, “American generosity is virtually unknown at home and abroad” (Pilon 2). This obsolete restriction “hampers public understanding of the U.S. government's activities” and impedes serious oversight of these programs (Pilon 4).

g) Develop a more culturally-based approach to public diplomacy

Cynthia Schneider and Mohan Dutta-Bergman, a professor of communication at Purdue University, argue that public diplomacy efforts need to focus more on cultural exchange. Schneider’s principal message is that “politics demonizes, art humanizes” (Schneider "Cultural Diplomacy: Hard to Define, but You'd Know It If You Saw It" 196). She underlines the potential of music, literature, and film to cross cultural boundaries and foster understanding and respect between countries and cultures. Schneider, like many other researchers in the field, references the Pew Global Attitudes Project as an indicator for the current global environment of anti-American sentiment and the need for improved cultural diplomacy programs (Schneider "Cultural Diplomacy: Hard to Define, but You'd Know It If You Saw It").

Like Peterson, Schneider is also concerned that current public diplomacy efforts sometimes undermine differences between the United States and other cultures. She argues that campaigns to discover common values between the United States and other cultures can obscure legitimate differences in perspective, and therefore show lack of cultural understanding, causing foreigners to receive our messages with resentment. The example she provides is the inability of U.S. public diplomacy officials to recognize the difference in roles given to families and individuals in the United States and in many Arab and Muslim countries. Schneider urges public diplomacy officials to proceed with caution when they emphasize the similarities between American and Arab or Muslim cultures (Schneider "Cultural Diplomacy: Hard to Define, but You'd Know It If You Saw It").
Mohan Dutta-Bergman also argues for a more culturally-based approach to U.S. public diplomacy efforts in the Middle East. Dutta-Bergman draws comparisons between current public diplomacy efforts in the region and similar efforts during the Cold War to demonstrate that our methods have largely remained the same in the past 50 years. Our public diplomacy efforts have been framed as one-way communication tactics that the U.S. government uses to influence and change the opinion of foreign publics, while maintaining its unwillingness to change U.S. opinions or policies. Dutta-Bergman critiques this one-way method, arguing that it generates further misunderstanding and resentment among foreign populations. He defends a more “culture-centered approach” which focuses on the relationships between cultures and establishes an ongoing dialogue centered on respect and the desire to develop mutual understanding. This system would contrast significantly with the current model of the United States as the “sender” of the message and the foreign publics as “receivers” who do not actively participate in the dialogue (Dutta-Bergman 119). Fundamentally, this new approach would shift the focus of public diplomacy from “informing and persuading to understanding, dialoguing, and relationship building” (Dutta-Bergman 121).

h) Increase oversight and improve evaluation techniques

Few of the existing reports on public diplomacy discuss direct measures of program efficacy, implying that this data is difficult to obtain or that the public diplomacy institutions do not currently have the resources to collect such data. Although public diplomacy institutions have recently increased oversight and evaluation of public diplomacy programs, the evaluation process still needs to be improved. During the 2007 hearing before the subcommittee on the Middle East and South Asia of the House Committee on Foreign Affairs, the committee members and the panelists raised many important issues concerning the efficacy of public
diplomacy programs in these regions of the world. The panel speaking before the subcommittee included representatives from various public diplomacy institutions (IIP, R/PPR, ECA, and the BBG). The panel gives insight into recently-implemented measurement methods that assess program impacts. They discuss many specific evaluations, polls and surveys that have been administered already ("Public Diplomacy in the Middle East and South Asia.").

At the subcommittee hearing, the panel also claimed that ECA has the most advanced “culture of measurement” of all the public diplomacy institutions. Other institutions are trying to follow the lead of those at ECA. They recognize the problem of quantifying the success of their programs and realize that it is important to find ways to measure program results and determine whether or not their programs are achieving their goals ("Public Diplomacy in the Middle East and South Asia."). The discussion of the panelists indicates that it is now possible to look at evaluations for specific programs and determine whether they are accomplishing their stated goals.

Problems with oversight and the evaluation process were particularly evident at the BBG and were very troubling to committee members. There were some major “mistakes made” in recent years at the BBG; for example, there were “accidental” broadcasts of a Holocaust-denier conference and a message delivered by a terrorist on Alhurra TV, which broadcasts throughout the Middle East. The committee was clearly concerned about the lack of oversight at the BBG. However, Mr. Blaya and Mr. Hirschberg of the BBG both claim that the proper measures are now in place to avoid committing such mistakes again ("Public Diplomacy in the Middle East and South Asia."). If clear evaluation procedures had already been outlined, then the BBG might have been more careful, for it would have been clear that these mistakes would have led to a rather negative evaluation of Alhurra, which could impact future program funding.
Many researchers have suggested that the BBG must improve their ability to measure the effects of their programs (Epstein and Mages 16). The GAO has recognized that the State Department has taken significant steps towards instituting more systematic evaluation practices. However, the GAO has critiqued the BBG’s use of audience size as a key performance measurement due to methodological concerns about how this number is calculated (Ford).

C) How the State Department and the BBG have changed since September 11, 2001

Since September 11th, there have been significant changes in the structure and content of public diplomacy programs. Notably, there has been substantial turnover in the office of Under Secretary for Public Diplomacy and Public Affairs, with multiple officials resigning after a year or two on the job. There have been six different Under Secretaries for Public Diplomacy and Public Affairs since the position was created in 1999 (Evelyn Lieberman, Charlotte Beers, Margaret Tutwiler, Karen Hughes, James Glassman, and Judith McHale) and each official has shown a somewhat different approach to public diplomacy programs, progressively incorporating more of the previously-mentioned recommendations into their efforts ("Under Secretaries of State for Public Diplomacy "). In May 2009, President Obama appointed a new Under Secretary, Judith McHale, the former President of Discovery Communications ("Biography: Judith A. McHale").

In 2007, the GAO noted that the State Department and the BBG had incorporated many of the recommendations of academics and government officials into their activities. Both institutions have established more precise strategic frameworks for their activities and developed more rigorous measures of effectiveness (Ford 3). However, the GAO also reported that the percentage of worldwide public diplomacy positions at the State Department that were vacant actually increased from 15% in 2006 to 22% in 2007(Ford 3). This problem is particularly
serious in the Muslim world, where language-designated public diplomacy positions continue to be filled by officers without the appropriate language proficiency, hampering their ability to engage foreign audiences. In response to this problem, in January 2006, President Bush launched the National Security Language Initiative to help increase critical foreign language training at all levels of education. Under the program, the State Department received increased funds to boost the language skills of Foreign Service officers. Secretary of State Condoleezza Rice also changed promotion criteria at the State Department to require more advanced language skills. These efforts may improve the foreign language deficit, but further improvement is still needed (Zeller 23). As of 2007, significant foreign language deficiencies were still a major problem, particularly in Arabic language posts (Ford 15).

Security concerns have also limited U.S. public diplomacy efforts in certain countries, where the “threat level for terrorism is rated as ‘critical’ or ‘high’ in 80 percent of posts.” However, in many of these predominantly Muslim countries, like Pakistan, these efforts are particularly important. As a result, embassies have had to close American Centers and find other, more secure venues for their programs and enforce strict security procedures for visitors, often sending the message that the United States is unapproachable or distrustful. The Department of State has attempted to deal with this issue by establishing “American Corners,” centers that provide information about the United States, throughout the world, particularly in the Muslim world (Ford 16-17).

Additionally, the State Department has increased the number of exchange students visiting the United States and facilitated the student visa process. Student participation in U.S. education and exchange programs grew from 27,000 to 39,000 between 2004 and 2007. In 2007, the State Department was planning to expand their programming to younger students, ages 8 to
14. The Under Secretary also established the Rapid Response Unit, which monitors international media and informs American policy makers with a daily report covering the events and ideas that are driving world news, and provides the U.S. position on those issues to an email list of thousands of senior officials (Hughes).

Former Under Secretary for Public Diplomacy and Public Affairs James Glassman, as well as his predecessor Karen Hughes, have both discussed recent efforts to expand programming. Both mention the Bureau of International Information Programs' (IIP) new digital outreach team, which counters misinformation and myths on Arabic-language blogs (Hughes; Glassman). Glassman continues to stress the importance of utilizing new technologies to help deliver our message abroad, like using Google or Facebook in new ways. He continues to emphasize the importance of framing the message and the influence of semantics on foreign audiences ("war of ideas" vs. "global ideological engagement"), as well as the continued difficulty of effectively measuring the results of public diplomacy programs (Glassman).

Christopher Midura, the Acting Director of the Office of Planning, Policy and Resources at the State Department (R/PPR), also discussed, in 2008, how the rest of public diplomacy is now operating under the same “culture of measurement” as ECA. He mentions the Evaluation and Measurement Unit (EMU) and the successful widespread use of The Mission Activity Tracker to evaluate public diplomacy programs. He also intends to increase investment in the work of EMU in the coming months (Midura).

President Bush also established a Policy Coordination Committee on Public Diplomacy and Strategic Communications in 2006, in an attempt to better coordinate interagency activities. However, there is still debate about the efficacy of this committee and whether or not there is effective interagency collaboration (Ford 9).
D) Next Steps for Public Diplomacy

Since the transition into the Obama administration, the GAO and academics like Michael Sandel have once again stressed the importance of public diplomacy efforts. Improving the United States’ image abroad is the number five priority on the GAO’s list of 13 urgent issues requiring Obama’s attention during the first year of his administration. “In today’s highly volatile global environment, it is more critical than ever that the United States effectively coordinate, manage, and implement its public diplomacy and strategic communications activities to affect foreign public opinion,” the GAO said (qtd. in Pincus).

In a *New York Times* article published shortly before Hillary Rodham Clinton became the new Secretary of State, various foreign policy experts include the most important questions they feel need to be addressed by the new Secretary ("Questions for Mrs. Clinton"). The questions of Michael Sandel, a professor of government at Harvard, highlight the continuing challenges facing public diplomacy efforts. He underscores many of the problems with our public diplomacy efforts already discussed by many of the previously-mentioned articles. For example, he asks whether we should drop the term “war on terror” and describe our policy more precisely as a war to defeat Al Qaeda and violent Islamic extremism. He also calls attention to the following paradox in the Middle East: countries with pro-American governments have populations with high levels of anti-American sentiment (Egypt, Saudi Arabia, Jordan) and countries whose governments are hostile to the United States have populations with more favorable views of the United States (Iran). How can we explain this trend and what does it say about our public diplomacy efforts? Sandel also states that one of the most damaging legacies of the Iraq war is that it has given idealism and internationalism a bad name. He asks Clinton how she will persuade the American people and the world that the United States can be a force for
democracy and freedom ("Questions for Mrs. Clinton"). These questions imply that, to this day, there are still a lot of problems with our public diplomacy efforts.

Nevertheless, although many critics cite public opinion polls like the Pew Global Attitudes Project as evidence for the inefficacy of public diplomacy programs, it is not actually possible to know whether or not these programs are accomplishing their goals without proper evaluation methods that tie measurable results directly to these programs. Many authors who have previously written on this topic seem to jump from broad characterizations of the problem (high levels of anti-American sentiment abroad) to policy recommendations without intervening analysis. As the ExpectMore.gov evaluation of Broadcasting in Arabic indicates, there are numerous factors - political, geographic, historical, educational, economic, ethnographic, diplomatic, and humanitarian - that influence public opinion in a given country, making it difficult to separate the impact of specific public diplomacy programs from this array of other factors (ExpectMore.gov "Detailed Information on the Broadcasting in Arabic Assessment"). As Foreign Service Officer and public diplomacy expert Joe Johnson accurately explains, “opinion polls by themselves set a standard that cannot be met, because those numbers go up and down for all kinds of reasons. And that is a problem for the practitioners. If you cannot define success, you’ll never succeed” (Johnson 44).

In fact, even if our public diplomacy efforts were somehow perfect, there would still be other factors that undoubtedly have more influence on foreign opinion of the United States than our public diplomacy programs. As Thomas Friedman aptly recognized in a New York Times op-ed shortly after Obama received the Democratic nomination, “it would not be an exaggeration to say that the Democrats’ nomination of Obama as their candidate for president has done more to improve America’s image abroad — an image dented by the Iraq war, President Bush’s
invocation of a post-9/11 “crusade,” Abu Ghraib, Guantánamo Bay and the xenophobic opposition to Dubai Ports World managing U.S. harbors — than the entire Bush public diplomacy effort for seven years” (Friedman). If Friedman is correct, that political factors can have a much bigger impact on public opinion than public diplomacy programs, this conclusion calls into question the impact that such programs can actually have. It is difficult to influence foreign populations with public diplomacy when there are so many other events and sources that also influence foreign opinion. Additionally, some experts question whether we can “temper the damage” when U.S. foreign policy is “fundamentally at odds with foreign public opinion” (Zeller 20). Nevertheless, these programs can still have an effect and it is still possible to find alternative methods to isolate the outcomes of U.S. public diplomacy programs.

The BBG, as well as ECA and IIP have recently taken steps to separate the influence of their programs from other factors. Because this system of measurement is relatively new in the public diplomacy arena, few scholars have looked at how public diplomacy programs are evaluated and what these evaluations demonstrate. It is necessary to examine these evaluation methods to determine whether IIP, ECA, and BBG programming are indeed accomplishing the overall goals of public diplomacy as well as the individual goals of each institution.

My evaluation will raise a variety of questions about the evaluation of public diplomacy programs. For example, although the three public diplomacy institutions contribute to American public diplomacy efforts, the programs administered by each have different audiences and different impacts. How do the differing scopes and goals of the programs affect the evaluation process? How do these institutions evaluate their own programs and do the questions they ask truly assess the goals of public diplomacy? Do some of these programs appear to be more effective than the others, or are they equally effective (or ineffective), but just in different ways?
III. Methodology


The primary framework for this analysis was a comparative case study of four program evaluations. All four evaluations were administered since September 11th, 2001 and all four programs are in the Middle East and South Asia. The method used was a qualitative meta-analysis of specific, existing evaluations to determine how well these programs are accomplishing their goals (or a “meta-evaluation,” an evaluation of evaluations) (Mark, Greene and Shaw 3). The goal of this study is to look for insights into the evaluation process.

The cases include evaluations of the following public diplomacy programs: 1) The English Access Microscholarship Program, 2) The BRIDGE (Building Respect through Internet Dialogue and Global Education) Program, 3) The International Visitor Program: Special Initiative Projects in Response to the Events of September 11, 2001, and 4) Alhurra Television. The primary reason for focusing on these programs is that they are all U.S.-funded public diplomacy programs in the Middle East and South Asia that have been administered since September 11, 2001, for which complete evaluations are currently available (they were the only programs to meet all of these criteria)\(^1\). The need to have complete evaluations does limit what I

---

\(^1\) While I have been unable to find any IIP evaluations on the internet, Sarah Brewer, Senior Evaluation Officer at the EMU of the State Department’s R/PPR did send me the executive summary of the evaluation of the American Corners Program (a program administered by IIP). Unfortunately, she was unable to find the full report since the reorganization of offices at the State Department has made it difficult to track down reports written prior to the dissolution of the PDEO (Brewer). The American Corners Program was not included in this evaluation because the executive summary alone does not provide enough insight into the evaluation process. However, including this program certainly could have been helpful, as it differs significantly from the other programs included in this analysis. “American Corners” are small libraries located throughout the world containing English-language books, magazines, periodicals, and multimedia materials with information about the U.S. These libraries often contain computer workstations and free Internet access. “American Corners” are designed to be easily accessible to the general public and are located in public and university libraries. There are also outreach programs, events and speakers hosted at the American Corners locations. The American Corners evaluation focuses on programs in Malaysia, South Korea, Indonesia, and Thailand. I would have been particularly interested in the results for Malaysia and Indonesia, two countries with substantial Muslim populations. This program influences a wider and broader participant-base than the ECA programs, although the primary visitors are educated young adults and college students. Additionally, the American Corners program involves less interpersonal interaction than the ECA.
will be able to conclude about public diplomacy programs generally, as these four programs are not necessarily “representative” of all public diplomacy programs. The selected programs do not just vary on one dimension, which may make it difficult to determine exactly which factor causes a difference in results between programs. However, while it may not be possible to pinpoint a specific cause for differing program outcomes, the program evaluations are still comparable. It is useful to compare the evaluations and the evaluation processes to learn which evaluation methods are most effective and accurate for measuring the impacts of each program.

a) The English Access Microscholarship Program (Access)

The English Access Microscholarship Program is an ECA program that provides an English language learning experience for 1-2 years, using an American-style teaching method, to non-elite (by standards of the country) 14 to 18 year-old students in countries with significant Muslim populations (Aguirre Division Evaluation of The English Access Microscholarship Program x). This evaluation specifically addresses the Access programs in Morocco, Lebanon, Oman, India, Bangladesh, and Pakistan (Aguirre Division Evaluation of Access Program ix).

b) The BRIDGE Program

Like the Access Program, the BRIDGE Program is also an ECA English-language training program for students in countries with large Muslim populations (Jordan, Lebanon, Morocco, and Pakistan). However, the BRIDGE Program primarily utilizes technology (computers, the internet, blogs etc.) to create an inter-cultural exchange with students in the United States, as well as students in BRIDGE Programs in other countries (Aguirre Division Evaluation of the School Connectivity Program (SCP) and the Building Respect through Internet Dialogue and Global Education Program (BRIDGE) 6-7). Unlike Access Program participants,

______________________________________________________________

programs, but is still probably more interactive than Alhurra (American Corners Program Pilot Evaluation Executive Summary Report). Brewer also informed me that EMU is currently working on other evaluations (Brewer).
the BRIDGE students have significant contact with students in the United States. Notably, the exchanges are mainly via the internet, not in person. The Access Program does not involve any exchange with U.S. citizens, except if the teacher happens to be American and during occasional visits by the U.S. ambassador or other U.S. public diplomacy officials.

c) The International Visitor Program (IVP)

The IVP, another program administered by the ECA, “brings current and emerging foreign leaders in government, media, politics, academia and other fields” to the United States for 3-4 weeks “to confer with professional counterparts and gain a deeper understanding of the United States and its institutions and culture” (Office of Policy and Evaluation executive summary, 2). The Special Initiative Projects focus specifically on leaders from countries with significant Muslim populations. The program included participants from Algeria, Egypt, Indonesia, Israel, Kazakhstan, Kyrgyzstan, Lebanon, Jordan, Malaysia, Morocco, Pakistan, the Philippines, Syria, Tajikistan, Thailand, Tunisia, Turkmenistan, Uzbekistan, the West Bank and Yemen. Unlike the Access and BRIDGE Programs, which both target high school students, IVP is clearly designed for an older age group. Additionally, the IVP is a physical exchange program in which the participants come to the United States, unlike the BRIDGE Program, which is largely an internet exchange program.

d) Alhurra Television

Alhurra (Arabic for “the free one”) television is substantially different from all three of the ECA programs because of the institution responsible for the program and the medium used. Alhurra is an Arabic television network, administered by the Broadcasting Board of Governors (BBG), as part of the Middle East Broadcast Network (MBN). The Alhurra television station is available in 22 countries in North Africa, the Near East and the Persian Gulf (Algeria, Bahrain,
Chad, Djibouti, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Palestinian Authority, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Yemen ("Alhurra TV"). Alhurra reaches a much wider and broader audience than any of the ECA programs. At the same time, the personal connections involved in the ECA programs are largely absent in Alhurra programming.

B) Program Evaluation Process

The structure of and techniques used in the evaluations vary from one program to another. Each evaluation uses a slightly different methodology; some combine multiple data measurement techniques, while others only use one. The types of organizations performing the analysis also differ. The time frame of the evaluation varies, along with the location of the programs. The reach and scope of the programs differ, as do the relevant sample sizes.

a) The English Access Microscholarship Program

In September 2005, the PDEO hired the Aguirre Division of JBS International, Inc., a private company that provides a broad and diverse set of management and information technology services, to conduct a pilot evaluation of the Access Program ("JBS International: Professional Services"). The field work was conducted in April and May 2006. The evaluation studied the programs in three countries in the Near Eastern Affairs (NEA) region: Morocco, Lebanon, and Oman, and three countries in the South and Central Asia (SCA) region: India, Bangladesh, and Pakistan. Data collection consisted of both quantitative and qualitative methods, including in-depth individual interviews, survey questionnaires, focus groups, discussion groups, and classroom observations. Data was gathered from 613 students, 137 peers, 136 parents, 35 teachers, and 21 administrators across 27 Access schools in these 6 countries. The evaluation also included interviews with Public Affairs Section staff and Regional English
Language Officers in select schools in each country (Aguirre Division Evaluation of Access Program xii). Although the evaluators do include some recommendations for program improvement, the evaluation is generally very positive. The evaluators conclude that the Access program is “highly effective” (Aguirre Division Evaluation of Access Program 93).

b) The BRIDGE Program

The BRIDGE Program evaluation was also conducted by the Aguirre Division of JBS International, hired by the Office of Policy and Evaluation at ECA (ECA/P). Because there are many countries included in the BRIDGE program (and the School Connectivity Program (SCP), which was evaluated simultaneously), Aguirre designed a methodology that combined select qualitative country case studies with quantitative surveys of program participants (Aguirre Division Evaluation of the SCP and BRIDGE Program viii). The evaluators chose to do a case study of 7 different schools involved in the BRIDGE program in Lebanon, where evaluators observed classes, interviewed staff, and conducted focus groups with program participants (Aguirre Division Evaluation of the SCP and BRIDGE Program 13, 16). Aguirre chose Lebanon for the BRIDGE case study because, while foreign language instruction in schools is common in Lebanon, the use of Internet-based curriculum and online collaborative-learning projects is not widespread. According to Aguirre, this provided a solid baseline against which to evaluate the impacts of the BRIDGE Program. BRIDGE Programs in three other countries (Jordan, Morocco, and Pakistan) were also represented in the online surveys of students, teachers and Master Trainers. In all of the BRIDGE countries, data was collected from a total of 101 students, 35 teachers and 5 trainers (Aguirre Division Evaluation of the SCP and BRIDGE Program 8, 17, 111). The sample sizes would be much bigger if the participants in the SCPs were included as well. However, I chose to focus on the BRIDGE Program exclusively
because of the countries participating; the countries involved in the related SCPs were generally not predominantly Muslim countries in South Asia or the Middle East. The JBS evaluators conclude that “there is substantial evidence of desired outcome achievement across the Programs,” an assessment that is perhaps not quite as positive as the Access assessment, but still quite favorable (Aguirre Division Evaluation of the SCP and BRIDGE Program 171). The evaluators do also recommend some changes to the Program based on their findings.

c) The International Visitor Program

The evaluation of the IVP was administered by the Office of Policy and Evaluation at the Bureau of Educational and Cultural Affairs (ECA/P), in consultation with the Office of International Visitors. The data was collected using a 19-question survey with both closed and open-ended questions, distributed to program participants at the end of their stay in the United States. Most participants completed the survey while still in the United States and a few completed it shortly after their program ended, back in their home countries. The evaluation contains results from 85 respondents out of a total of 95 participants, from 10 different IV Special Initiative projects and 19 different countries in the Near East, South Asia, East Asia, and Central Asia. The report represents the aggregated results from these 10 programs administered between February and October 2002. The sample included representatives of foreign ministries and ministries of defense, leaders of non-governmental organizations and local government, journalists, community and civic leaders, national government officials, and student leaders in educational institutions (Office of Policy and Evaluation 1, 4-5). Since this evaluation only used one data collection technique, it is less comprehensive than the Access and BRIDGE evaluations. Notably, the IVP evaluation was also not conducted by an outside contractor like Aguirre, which may influence the evaluation process. This evaluation was also very favorable; the evaluators
claim that the IVP was “very successful in providing accurate and balanced information about the United States, and “significantly contributed to ECA’s counter-terrorism goals” (Office of Policy and Evaluation 16). While the evaluators do discuss some recommendations for program improvement, they do not discuss these recommendations in nearly as much detail as do the JBS evaluators of the Access and BRIDGE programs.

d) Alhurra television

The Alhurra evaluation was conducted for the BBG by The Center on Public Diplomacy at the Annenberg School at the University of Southern California in 2008. The evaluation was based on content analysis of 77 hours of Alhurra programming during November 2007, as well as expert discussion group proceedings. The shows were categorized by type and unitized for analysis. A code-book containing the rules to be used in the coding process was written and native Arabic-speaking coders were trained according to these rules. The expert discussion groups took place in Beirut, Cairo, and Dubai and were comprised of academics and news professionals from their respective regions. The participants had watched Alhurra for at least one month before the discussions, which took place in March and April 2008. They were shown representative stories from the November 2007 programs used in the content analysis. There were 12 participants in Beirut, 19 in Cairo, and 11 in Dubai (USC Center on Public Diplomacy 11, 70-74). Unlike the other three evaluations, the Alhurra evaluation was quite negative. The evaluators state that “Alhurra is not performing at the level that it needs to reach to be successful” (USC Center on Public Diplomacy 4).

C) Analysis of Program Evaluations

These program evaluations will be analyzed systematically using specific criteria, drawing on the literature about qualitative and quantitative program evaluation and performance
measurement. The chosen criteria are: a) the validity and reliability of performance measures, b) the degree to which the evaluation is statistically sound, c) the strength of causal inference and the interpretation of the data, d) the relative neutrality of the evaluators and e) the examination of the program in a broader geographical, social, and cultural context. These criteria will be used to critically analyze each evaluation and determine the extent to which the claims of each evaluation can be supported.

a) The validity and reliability of performance measures

It is essential to consider the validity and reliability of the chosen performance measures in each evaluation. Initially, to measure performance, evaluators must provide a clear definition of the objectives and purpose of the program at hand (Moore 159). There must also be recognized standards for judging whether or not programs are successful. The evaluators must clearly state the criteria they use to determine the effectiveness or success of a program for the evaluation to be useful.

To be valid, a performance measure must reflect the actual program outcome that the evaluators claim it reflects. According to the 1993 Government Performance Results Act (GPRA), federal agencies should choose performance measures that “1) are closely aligned with their stated goals; 2) approximate actual performance as closely as possible; 3) are relatively simple and inexpensive to administer; and 4) make it difficult for managers to increase their measured performance in ways other than increasing their actual performance” (Heinrich 714). The evaluators should make a convincing argument that the performance measures used are closely related to program impacts (Heinrich 717). If the evaluators use many different performance measures, they should indicate which ones are most important and indicative of program effectiveness (Nyhan and Marlowe Jr. 336).
The performance measures should also be reliable; if someone else were to conduct the same analysis using the same performance measures, they should theoretically obtain the same results. It is nearly impossible to replicate a study exactly, since the time frame of the study will always impact the results to an extent. However, an evaluation will still be considered reliable if it can be replicated by another evaluator who would otherwise receive the same results, independent of the time frame of the evaluation.

b) The degree to which the evaluation is statistically sound

This analysis will also address the typical statistical issues that come up in any evaluation, such as sample size, biased responses, and selection bias. Are the samples big enough to draw relevant conclusions? Generally, the larger the sample size, the more confidence one can have that the results accurately reflect what is occurring in the population. Small samples can adequately represent homogenous populations, but larger samples are necessary for more heterogeneous populations (more elements are needed to represent all of the diversity). However, while it is generally true that larger samples will be more representative of the population than smaller ones, there is a point of diminishing returns for increasing sample size. Doubling the sample size will double the cost of the survey, but will increase the accuracy of the data by significantly less than double. Therefore, researchers tend to use sampling ratios to determine acceptable sample sizes. As a general rule, the smaller the population, the larger the sampling ratio needs to be (Ruane 108-9).

In addition to sample size, it is important to determine whether the phrasing of the survey and interview questions is objective or is likely to elicit biased responses. Additionally, since participants were not randomly assigned to take part in any of the selected programs, selection bias may potentially be a problem. The samples are not randomly selected because certain
criteria were used to decide who would participate in the program (people had to apply and show interest in the program, they needed certain professional or academic qualifications or needed to demonstrate that they were economically disadvantaged). If the samples had been randomly-selected, the researchers would have randomly decided which members of the given population would participate in the program and which would not; they would not use other criteria involving the individual’s abilities or circumstances to determine participation. Without random assignment to groups (participants and non-participants), it is difficult to determine whether the observed difference between the groups is a result of the treatment or of initial differences between the two groups. In other words, the participants and non-participants may be systematically different from each other initially, regardless of the effects of the program itself.

c) The strength of causal inference and the interpretation of the data

Even if researchers properly collect the data and use the appropriate statistical tools, they will not necessarily make sound, rational policy recommendations based on the data. In part, this is a problem of causal inference. Good evaluations appropriately consider questions of causality and control for confounding variables and bias. When evaluations are used in policy-making, cause-effect statements are inherent; the evaluator must estimate the effects of the program (compared with those of policy alternatives) to determine the contribution of the specific program. Causal methods can also be used in evaluation simply for the purpose of knowledge development and can facilitate learning about the causes of problems and their potential solutions. Additionally, causal methods may be used to determine how a program has its effects. It is helpful to understand whether or not a program works, but it is even more instructive for policy purposes to understand how and why it works. Causal methods can also be useful in
determining the significance of a problem and may reveal whether there is a need for greater social intervention (Mark and Henry 318-20).

When examining the question of causality, evaluators should address the possible existence of mediator and moderator variables and try to control for these variables. A mediator is a variable that falls between a program and its outcome in a causal chain and is responsible for the relationship between the treatment and the outcome. Before determining the effects of a program itself, it is important to rule out any mediator variables. In contrast, a moderator modifies the strength or direction of the relationship between the program and the outcome; it is associated with stronger or weaker effects of the program on the outcome (Mark and Henry 330-1). As an example, consider the possible relationship between participation in an educational exchange program and favorable attitudes towards Americans. Education might be a mediator variable in that it may explain a theoretical relationship between participation and favorable views of Americans. It may be the case that more educated people tend to be more open-minded about Americans and more likely to participate in these programs. If you remove the effect of education and the relationship between participation and favorable attitudes towards Americans disappears, then education is a mediator variable. The participant’s parents’ views of Americans may be a moderator variable, in that the relationship between participation and favorable attitude towards Americans could be stronger for participants whose parents have relatively favorable views of Americans and weaker for participants whose parents have strong negative impressions of Americans.

Furthermore, it is difficult to demonstrate causality when subjects are not randomly assigned to groups. To address these difficulties, some evaluators may use quasi-experiments. Like randomized experiments, quasi-experiments involve comparisons across multiple
conditions, but they lack random assignment to conditions. Quasi-experiments can be used to inform policy-making when randomization is infeasible for practical or ethical reasons. Of course, if a quasi-experiment is used, it becomes difficult to rule out selection bias. To limit this difficulty, an evaluator could also do a pretest to see whether participants and non-participants were systematically different before the program was administered (Mark and Henry 323-24).

d) The relative neutrality of the evaluators

When analyzing these evaluations, one must consider who is doing the evaluation and what their stake is in the evaluation results. Ideally, program managers want to improve the quality of their organization and will not try to manipulate the evaluation system simply to promote budget increases. Unfortunately, this ideal situation often does not match reality (Nyhan and Marlowe Jr. 345). Furthermore, there are traditionally several different types of organizations that conduct evaluations, including auditing agencies, contract evaluators hired by governments, other independent institutions such as universities or other private and public agencies, or internal evaluation units of governments. Each of these institutions has something different to contribute to the evaluation process and none can respond to all of the broad and specific needs of an evaluation of a given program. Different types of evaluators may use distinct methods or approaches in evaluations.

Auditing agencies usually operate with a general mandate and randomly decide which activities to audit, making them unable to audit all activities systematically and continuously. Outside contractors who conduct evaluation work for the government are generally assumed to be independent from government influence. However, these contractors are funded by the government and are often hoping to be rehired in the future. Therefore, these evaluators may compromise their objectivity to ensure continued contracts. Independent institutions often do not
focus on issues of importance for government interventions because independent institutions have the privilege and discretion to choose and prioritize the objectives of the evaluation (Davies, Newcomer and Soydan 172, 74).

Because of the drawbacks of these non-governmental evaluators, governments have increasingly established their own units of evaluation to secure evaluation in specific policy areas. However, there are problems with government management of evaluation as well. For example, there may be reason to question the legitimacy and credibility of an evaluation if the evaluators are funded from the same agency budget that funds the evaluated programs. Political considerations could also lead the government to only publicize those evaluations that are supportive of the programs favored by the government. Thus, the credibility of the assessment process may be undermined. Additionally, one might question the transparency of governmental evaluation (Davies, Newcomer and Soydan 172-4).

e) The examination of the program in a broader geographical, social, and cultural context

It will also be helpful to address whether the evaluators look at the given program in a broader context versus in isolation. Simply because a program has a particular outcome on a specific group of participants in a specific region does not mean that a similar program (or even an identical program) will have the same impact on a different community. Therefore, an evaluator should recognize the impact of the program setting on its success or failure, and should not simply look at the program independent from the context in which it is administered.

The broader context of a program encompasses several different issues. First, this broader context includes whether the evaluators take into account the geographical location of the program at hand. It may be the case that a program that functions in one country or city, with a certain set of cultural, social, and religious norms, could work very differently in another
region. Additionally, do the evaluators consider how the specific program may differ from other programs of its kind? If the evaluators attempt to generalize the results of their evaluation to other similar programs, they should make a convincing case that the other programs are in fact comparable with the program being evaluated. Evaluators who examine the program in a broader context would also consider the effects of the program on people other than the program participants, like participants’ family, friends, and the wider community.

Furthermore, the time frame for the evaluation is also an important part of the larger context in which the program was evaluated. Were the evaluations distributed during, immediately after, or significantly after the program? Do the evaluators distinguish between short-term and long-term effects of the program and recognize that the time frame for their analysis could greatly impact their results? It is possible that participants could have significantly different reactions two days after the program ended versus two years after the program ended. Evaluators should be conscious of this possibility.

If much of the analysis relies on the participants’ answers, do the evaluators look at results from other similar programs to decide whether participants’ responses are reliable? For instance, if participants say they will get more involved in the community, is there evidence from similar programs that these participants do actually get more involved? The extent to which the evaluators address these questions about the context of the program and of the program evaluation will be important in determining the strength of the analysis.

D) Limitations of this research

One limitation of this research, which is a common limitation of any research project involving case studies, is the ability to generalize my findings to the field of public diplomacy in general. The ultimate objective of this research is to analyze these select evaluations and make
statements about the evaluation process and about public diplomacy programs as a whole. The goal was to include a range of programs so that the findings of this research could be generalized. However, there will still be some public diplomacy programs that do not fall neatly into one category of programming or another. Furthermore, while my analysis of the evaluations may lead to insights about the programs themselves, I am still primarily looking at the validity and reliability of the evaluation process and only have limited ability to make judgments about the true efficacy of the programs.

Additionally, the United States administers public diplomacy programs throughout the entire world, and many of the programs outside of the Middle East and South Asia are essential, given that favorability of the United States among traditional allies (Germany, Britain, France, Spain, Canada) has also decreased in recent years (The Pew Global Attitudes Project: Global Unease With Major World Powers Rising Environmental Concern in 47-Nation Survey 4-5, 15-16). Additionally, support from allies in Canada and Western Europe would arguably make our public diplomacy efforts in the Muslim or Arab world more successful. However, researching programs throughout the entire world is beyond the scope of this paper, which will focus solely on efforts in predominantly Muslim countries in the Middle East and South Asia.
IV. Empirical analysis: Applying these criteria to evaluations of public diplomacy programs

I examined the four evaluations using the aforementioned criteria. This section will discuss the results of that analysis, detailing how each of the four programs performed according to each metric. I will discuss each criterion outlined in the methodology, for all four programs, and then continue to the next criterion. I discuss the validity and reliability of performance measures first because this is a fundamental concept for my evaluation, and provides a broad background on the types of questions that the evaluators asked in each evaluation. Because of the importance and breadth of this section, it will be longer than the others.

I will then discuss the next three metrics, 1) the degree to which the evaluation is statistically sound, 2) the strength of causal inference and the interpretation of data, and 3) the relative neutrality of the evaluators. These three metrics are actually just more specific indicators of validity and reliability. Next, I will discuss the fifth metric, the examination of the program in a broader geographical, social, and cultural context. This section is different from the others because it is less about the validity and reliability of specific performance measures and more about how the program relates to the program setting and other programs of its kind, and how the evaluators assess the overall impact of the program.

A) The validity and reliability of performance measures

a) Are the objectives of the program clear and are the standards for judging whether or not it is successful also clear?
In general, the three ECA programs have relatively well-defined goals. However, the IVP goals are not as clearly-stated as those of the Access and BRIDGE programs. Furthermore, Alhurra seems to have conflicting goals.

The objectives of the Access program are as follows: to provide a significant English language learning experience to non-elite students in underserved areas in a cost effective manner, to increase students’ ability to successfully participate in the socio-economic development of their countries and improve their chances of participation in future U.S. educational and exchange programs, and to provide students with an increased appreciation for American culture and democratic values (Aguirre Division Evaluation of Access Program x, 2). Based on the performance measures chosen by the evaluators, it is implied that the program is successful if students are proud of their English abilities, if they can get better jobs and apply to college because of what they learned in the Access program, and if students become more knowledgeable about the United States and about democratic values.

The evaluators also imply that program success depends on student and parent recognition that the program is sponsored by the United States, suggesting that another unstated goal of the program is to foster favorable views of the United States (Aguirre Division Evaluation of Access Program xiii-xiv). The evaluators state that the program is a “high profile, public diplomacy initiative that provides participating U.S. Embassies and Consulates an opportunity to demonstrate tangibly our country’s dedication to education and mutual understanding” (Aguirre Division Evaluation of Access Program 4). The evaluators also suggest that the students’ ability to meet with prominent U.S. political figures and representatives of the U.S. Embassy contributes to the success of the program (Aguirre Division Evaluation of Access Program 22).
The BRIDGE program evaluators also lay out the program goals clearly. As would be expected, the objectives are rather similar to those of the Access program, but with more emphasis on technology use. The program is supposed to: 1. Build understanding between U.S. citizens/communities and citizens/communities overseas participating in the program, and among countries, regionally and internationally; 2. Catalyze and assist educational reforms, including project-based learning, collaborative learning, and using technology in the classroom; 3. Develop skills and knowledge for teachers, students, regional coordinators, trainers, and other participants; 4. Support and enhance English instruction, and 5. Build sustainability in terms of resources, activities, and linkages (Aguirre Division Evaluation of the SCP and BRIDGE Program 6-7). If the program accomplishes these goals, it is deemed successful.

The objectives of the IVP are not as clearly-delineated. The evaluators initially state that the primary goal is to “increase understanding through dialogue at the personal and professional levels” by “creating opportunities for foreign visitors to meet and confer with their professional counterparts in the United States.” The Special Initiatives Projects are intended to engage emerging leaders from countries with large Muslim populations, promote understanding, and counter misperceptions of and negative attitudes towards the United States (Office of Policy and Evaluation 2). The evaluators then elaborate that the primary goals of the evaluation were to assess the counter-terrorism objectives of the IVP, which include providing accurate information about the United States, deepening foreign understanding of the United States and U.S. goals, sharing information and establishing dialogue with moderates in Muslim countries, and broadening mutual understanding between the United States and predominantly Muslim countries. The emphasis on these counter-terrorism objectives suggests that these are the most important goals of the program.
The evaluators list the *secondary* goals of the evaluation as assessing changes experienced by the program participants (changes in personal views and perceptions, in learning and knowledge, in understanding of the United States). Implicitly, these changes are less important than the counter-terrorism goals, since the evaluators place less importance on them (Office of Policy and Evaluation 3). However, it is hard to understand how these goals could be secondary, for it is difficult to assess whether the counter-terrorism goals have been accomplished if we do not know how the views of the participants have changed. For instance, how can one appraise whether the program has deepened foreign understanding of the United States without assessing changes in understanding? How do we know if the program provided accurate information about the United States without asking about changes in knowledge of the United States?

The objectives of Alhurra, as laid out in the evaluation, are somewhat problematic. On the one hand, as the evaluators discuss, Alhurra must fulfill its responsibilities deriving from the Middle East Broadcasting Network (MBN) Journalistic Code of Ethics and from the International Broadcasting Act of 1994. These responsibilities include: providing comprehensive coverage, showing discipline in producing news without the influence of personal and institutional biases, offering diverse viewpoints, avoiding rumor, and thoughtfully and thoroughly addressing topics of greatest interest to the target audience (USC Center on Public Diplomacy 3).

Furthermore, Alhurra, like all U.S. international broadcasting, should be “consistent with the broad foreign policy objectives of the U.S.,” should “not duplicate the activities of private U.S. broadcasters” or the “activities of government supported broadcasting entities of other democratic nations,” and should meet needs which “remain un-served by the totality of media
voices available to the people of certain nations” (USC Center on Public Diplomacy 9). Alhurra is successful if the programming is accurate, balanced, comprehensive, and objective. Furthermore, Alhurra should serve as a “model of the free marketplace of ideas and a free press in the American tradition,” promote freedom, democracy and human rights, be independent from any political party or special-interest organization, and not endorse or advocate any specific political, economic or religious viewpoint (USC Center on Public Diplomacy 8-9).

At the same time, Alhurra is also supposed to “reflect and promote U.S. policies” (USC Center on Public Diplomacy 3). As I will discuss later in this analysis, this goal is difficult to accomplish when Alhurra simultaneously strives to be an objective and reliable news source.

b) **Are the performance measures valid?**

One of the first questions to ask when looking at these program evaluations is whether the performance measures are valid and reflect actual program outcomes. Comparing the types of questions asked in the evaluations to specific program goals can provide evidence of the relative validity of the measures. The results in this category differ somewhat by program. All four programs have at least some valid performance measures, as well as problems regarding the validity of other performance measures.

**Access**

The Access evaluators do use some valid performance measures in their evaluation, while other measures are undoubtedly problematic. The Access Program evaluators used various data collection techniques from a variety of sources, in an attempt to establish a balanced assessment of the program. The evaluators administered interviews, surveys, and focus groups, observed classrooms, and included the assessments of participants, peers, parents, teachers, and program administrators in their evaluation.
Participants were asked to assess their English skills, an important question given that one of the primary programmatic goals was to improve English language skills. The students rated their English language skills before and after the program; in all countries, participants reported a dramatic improvement in English language skills (Aguirre Division Evaluation of Access Program 63, 68, 72, 82, 85, 89). However, this measure does not prove that the program improved the students’ English skills; this is simply the students’ own assessment of how their skills have changed. Presumably, they expected that their skills would improve after devoting so much time and energy to the Access Program. It would have been helpful if parents and teachers were asked to assess the participants’ English skills before and after the program. Although the evaluators asked parents and peers if their own English skills improved because of the participant’s involvement in the program, they were not asked to assess the participants’ skills. Teachers also commented on the importance of English language training for economic advancement, but they did not directly assess the students’ skills before and after the program.

Access participants were also compared to a group of peers who did not participate in the program. Both groups were asked to assess their self-confidence, their views of the United States, their agreement with various democratic values, and their leadership skills. Participants generally had more favorable views of the United States and were more self-confident (Evaluation of Access Program 57-59). Peers, when compared with Access students, were more likely to rank their leadership skills as poor (Aguirre Division Evaluation of Access Program 62). Nevertheless, these differences do not necessarily prove that the program was the sole reason why participants had more favorable views of the United States or were more self-confident. The two groups may differ in other significant ways; the type of students who participate in the program may generally be more open-minded, ambitious, and confident than many of their peers.
Additionally, the difference in leadership skills is small and for some leadership skills, the proportion of peers who ranked themselves as excellent or good is higher than the proportion of participants who chose those rankings. Therefore, these measures do not actually tell us much about the program and the evaluators manipulate these measures to prove program success. Measures that are more useful are “changes in views of U.S. government and American people” and “changes in participants’ understanding of knowledge of the United States” (Aguirre Division Evaluation of Access Program 58, 60). At the same time, these measures still do not necessarily isolate the effects of the program.

The evaluators of the Access program do attempt to quantify the value of “multiplier effects,” “the impacts that program activities (or investments) have as they circulate through the program, its immediate environment (human, organizational or geographic), and into the community (a ripple effect)” (Aguirre Division Evaluation of the SCP and BRIDGE Program 154). In other words, multiplier effects are the impacts of the program on non-participants like parents and peers. For example, peers were asked whether they would be interested in taking Access classes, and all but one of the peers responded positively (Aguirre Division Evaluation of Access Program 56). Parents also reported that their Access students shared the knowledge they learned with them, with their siblings, and with other family members. 83% of parents said that seeing their children improve their English has inspired them to learn English. Furthermore, 72% of Access students have taught English to their parents or family members (Aguirre Division Evaluation of Access Program 99). Of course, these measures are not perfect either; simply because peers say they are interested in the program does not actually mean that they will become future participants and simply because parents are inspired to learn English right after their children complete the program does not necessarily mean they will actually learn English.
The evaluators also looked at the awareness of U.S. government sponsorship for the program. In most schools, students were aware that the program was funded by the U.S. government; however, there were some students, teachers, and administrators who misreported that the program was provided by the in-country educational service provider or by the government of their country. By including this measure in the analysis, the program evaluators were indicating that a well-established knowledge of U.S. funding is a program goal (Aguirre Division Evaluation of Access Program 98).

**BRIDGE**

The performance measures used in the BRIDGE program evaluation are similar to those used in the Access evaluation, given the similar goals of the two programs. The evaluators used focus groups, classroom observations, and online surveys to assess program performance. During the classroom observations and interviews with teachers and students in Lebanon, the evaluators assessed the following program goals: enhanced levels of mutual understanding, educational reforms, and English language instruction, acquisition and application. They also evaluated the students’ improvement in areas like computer skills, self-esteem, and teamwork, and the teachers’ emphasis on community involvement and social responsibility.

This part of the evaluation is qualitative and contains quotes from various teachers and students regarding their opinions on these issues (whether/how much the program enhanced mutual understanding, how much their English improved as a result of the program, changes that teachers noted in their students etc.). The quotes included in the evaluation are overwhelmingly positive and indicate that the program is accomplishing the aforementioned goals. Students and teachers alike see improvements in students’ self-esteem and their ability to work well in teams, and students have learned about the importance of community involvement and social
responsibility. However, because the quotes are not attributed to specific people, but rather, simply to a “student in Lebanon” or a “teacher in Lebanon,” the reader does not actually know how many different people these quotes represent or any characteristics of the respondents. It is also unclear whether the evaluators excluded any quotes that were unfavorable to the program (Aguirre Division Evaluation of the SCP and BRIDGE Program 49-54).

Teachers from one U.S. high school that participated in the BRIDGE program were also included in the qualitative evaluation. The teachers discussed program activities, including a physical exchange component; the U.S. students hosted several Jordanian students and one teacher and several students also travelled to Jordan. The lead BRIDGE teacher explained how her students and the students’ parents were initially hesitant to host students from the Middle East. However, their attitudes quickly changed after the visitors’ arrival; the American students developed ongoing friendships with the foreign students. Teachers explained how much they learned about other cultures through physical exchange; they developed a new level of respect for and understanding of the other culture. The evaluation only includes the responses of a few teachers, so we cannot truly judge the impact of the programs on the American students. Additionally, this is a very limited sample, as it is only representative of one particular program at one high school in Washington D.C (Aguirre Division Evaluation of the SCP and BRIDGE Program 55-57).

These qualitative measures do provide background information about how participants have responded to the BRIDGE program. However, it is important to look at the quantitative measures and see how they compare. At the same time, the two types of measurements are not fully comparable, since the qualitative data was only from students and teachers in Lebanon, while the online surveys include responses from students and teachers from Lebanon, Morocco,
Jordan, and Pakistan (Aguirre Division Evaluation of the SCP and BRIDGE Program 17, 111).

In general, the performance measures used on the online surveys are fairly strong because the evaluators ask for both the current level of knowledge, understanding, and skills and an assessment of change, in an attempt to single out the effect of the program on that particular piece of knowledge or skill. However, again, a pre-program and post-program evaluation would have been an even better measurement tool to isolate the effects of the program.

Students were asked the extent to which working on a joint project with a partner school in another country (often in the United States) helped them gain a better understanding of the partner’s culture or society. They were also asked the extent to which their participation in the BRIDGE program changed their understanding of U.S. values and culture and of daily life in the United States (Aguirre Division Evaluation of the SCP and BRIDGE Program 114-16). Students also assessed their computer and English language skills, and the extent to which their improved English skills have improved their performance in other classes (Aguirre Division Evaluation of the SCP and BRIDGE Program 115-21). Notably, some of these questions are rather subjective (the assessment of the United States more so than the assessment of specific skills). It may be difficult for students to assess how much the BRIDGE program alone has affected their understanding of the United States. Alternatively, they may feel somewhat pressured to say that the program has improved their understanding of the United States, since the evaluation is handled by a U.S. company and the program is funded by the American government.

Students were also asked to report the types of activities they participated in (online forums, website development, special presentations, research projects etc.) and the topics they discussed during program activities (conflict resolution, international or current events, English, American studies, gender issues, health etc). The evaluators note that very few BRIDGE
students reported discussing “American studies or holidays.” By calling attention to this statistic (as opposed to many of the others), the evaluators imply that it is problematic or at least, contrary to our expectations, that BRIDGE students did not learn more about U.S. culture (Aguirre Division Evaluation of the SCP and BRIDGE Program 112-13).

The BRIDGE teachers and master trainers were also asked about the activities and discussion topics they incorporated into their program, their interaction with other teachers within their country, the skills they learned from the program, and how their understanding of the United States and of Americans changed as a result of the program. They assessed how the BRIDGE program impacted English language training at their schools in general. These performance measures are generally similar to those used on the student surveys and come with many of the same concerns.

There is also an entire section of the evaluation devoted to multiplier effects. The evaluators asked program participants about the skills they shared with others; this included teachers sharing with other teachers, teachers sharing with other schools, students sharing with other students, students sharing with friends and family members. Next, teachers and students were asked whether they had organized or participated in community service projects, whether they have increased their volunteer activities, and if they will participate in community service activities in the future. It would probably be helpful to re-evaluate a few years from now and see if the students actually participate in more community service in the future (Aguirre Division Evaluation of the SCP and BRIDGE Program 156-57).

**IVP**
The performance measures used in the IVP evaluation have many of the same strengths and weaknesses as those used in the Access and BRIDGE evaluations. However, the IVP evaluation is based on only one research method, a survey of program participants.

The participants rated how valuable different program activities were, providing information to the program directors about which activities were most important or effective. Nonetheless, this question is only helpful to an extent. While the participants rank the relative value of different activities, they do not answer why these programs are valuable and how the value of these activities relate to the objectives of the IVP. The participants were also asked about the opportunities they had to share information with Americans on a range of topics, including information that related to their area of professional interest, and whether or not they were able to better inform Americans about their home countries. These questions attempt to evaluate whether the program increased mutual understanding and fostered the exchange of ideas between Americans and foreigners from various Middle Eastern countries. Participants also assessed how the program increased their understanding of various themes, all relating to U.S. society, politics, culture, or values. Since one of the programmatic goals is to provide accurate and balanced information about the United States, these measures are important (Office of Policy and Evaluation Appendix B, 1-8).

The respondents were also asked the extent to which they agreed or disagreed with certain statements about Americans. This question was clearly designed to demonstrate how the participants viewed Americans and to see whether certain negative misperceptions were still prevalent. However, this is a rather weak measure because the participants were not asked this same question before participating in the program nor does the question discern how participation in the program affected their opinions on the various statements. Therefore, simply
because participants may hold positive views of Americans, does not mean that the program changed their views. In fact, many of the participants in this program, who tend to be moderate, well-educated leaders in their respective countries, may have somewhat objective, moderate views of the United States to begin with (Office of Policy and Evaluation Appendix B, 8-9).

A question which better addresses the program goals, which is also included in the survey, is, “to what extent do you think your IV Program experience helped to (or will help to): a. promote closer relations between the U.S. and your country or region, b) strengthen ties between people of the U.S. and people of your country, c) promote mutual understanding and mutual respect among peoples of the world, d) provide you with a more informed and broader perspective of the U.S., e) enhance your understanding of diversity of opinion in the U.S., f) enhance your understanding of international issues in general, and g) develop an interest in professional collaboration with people you met in the U.S? (Office of Policy and Evaluation Appendix B, 9) The participants were also asked if they intended to share the information they gained when they return home (the overwhelming majority said yes) (Office of Policy and Evaluation Appendix B, 4)

At the same time, while the preceding two questions do address the major goals of the program, they are problematic because they assess the predicted long-term outcomes of the program, instead of the actual outcomes. There may be a significant difference between what the respondents say immediately after participating in the program and what they say and do a few months or years later. The evaluators do, at least, recognize this problem; they state that, “the evaluation results are based on immediate assessments…Therefore, it is recommended that a follow-up study be conducted in the future with the same participants…to learn if these
assessments…are sustained over time and to learn whether the International Visitors proceeded with their planned future activities” (Office of Policy and Evaluation 16).

Another downfall of this assessment is that the views of the Americans hosts are not included. We do not know how the Americans felt the program influenced their understanding of and views towards other cultures. Given that “the ultimate goal of any exchange opportunity is that ‘both sides’ benefit from the exchange of new information and the sharing of diverse perspectives” this evaluation only looks at half of the equation (Office of Policy and Evaluation 11). This begs the question: is the goal actually more one-sided than the program administrators care to admit, if the evaluators do not bother to survey American participants? Conversely, is the impact on American participants important to program administers, and the lack of data on the American hosts just reflects an error on the part of the evaluators?

Furthermore, the evaluators chose to highlight certain measures within the evaluation and gloss over others, indicating what they think are the most important statistics gained from the survey. For example, the evaluators claim that “participants universally reported an increase in understanding and changes in their perceptions of the United States in general, the government and policies of the United States, and other aspects related to life in the United States and American values, beliefs, and ways of living” (Office of Policy and Evaluation 7). While this is all true, what the evaluators fail to mention is that this increase in understanding was much smaller for “U.S. foreign policy,” and the “United States’ bilateral relationship with your country” (Office of Policy and Evaluation 7). These topics are two of the most important issues in the survey since one of the main goals of the IVP is to provide accurate information about U.S. policies and to foster positive relationships between the United States and other countries.

Alhurra
The Alhurra TV evaluators use strong performance measures; however, the conflicting goals of the program are evident. The evaluation includes data from both content analysis, performed by native Arabic-speaking coders, of all Alhurra news and topical programming during November 2007 and expert discussion groups conducted in Beirut, Cairo, and Dubai. The performance measures used to assess the TV content were: 1. The overall accuracy and comprehensiveness of the news agenda; 2. The avoidance of personal judgments and the use of value-neutral language; 3. The presentation of opposing views when covering major issues; 4. The avoidance of the presentation of unsubstantiated information; 5. The ability to provide a balance of sources in reporting the news; 6. Alhurra’s ability to clearly and effectively present U.S. policy in the region; 7. Alhurra’s treatment of religion and religious issues (USC Center on Public Diplomacy 13). The questions, and therefore the performance measures, that the coders addressed were essentially the same as those posed to the focus group participants, (USC Center on Public Diplomacy 38-47).

These appear to be valid performance measures; they all directly address the stated goals of the program, with the exception of number seven. However, Alhurra’s treatment of religion and religious issues is indirectly related to the program goals since one of the aims of the program is to promote freedom, including freedom of religion (USC Center on Public Diplomacy 58). Additionally, the discussion of religious issues could relate to other programmatic goals, such as impartiality, since religion is a particularly sensitive, sometimes controversial issue in many of the Arabic-speaking countries where Alhurra broadcasts (USC Center on Public Diplomacy 58). Furthermore, religion is likely to be a topic of interest for the target audience (USC Center on Public Diplomacy 3).
There is, however, one potential problem with the sixth metric, Alhurra’s ability to clearly and effectively present U.S. policy in the region. The evaluators imply that because Alhurra was more likely to promote U.S. policy than to describe it neutrally, that the reporters were not clearly presenting U.S. policy (USC Center on Public Diplomacy 33). While it is true that Alhurra strives to be an impartial news source, another goal of Alhurra is to “reflect and promote U.S. policies” (USC Center on Public Diplomacy 3). At the same time, the evaluators show that if viewers see Alhurra as American propaganda, they are less likely to view it as an objective news source (USC Center on Public Diplomacy 36). Thus, the evaluator’s criteria and the goals of the program appear to be in conflict, highlighting a possible problem with the evaluation, but also with Alhurra’s contradictory goals. Furthermore, the evaluators base their assumptions about viewer perceptions of the program on the viewpoints of the focus group participants; they do not have any broader measures of actual audience perception.

c) **Are the performance measures reliable?**

The performance measures used in both the Access and BRIDGE evaluations are only somewhat reliable measures; they are not likely to all produce the exact same results if someone else were to replicate the evaluations. The majority of the data came from surveys and interviews with the students and teachers and presumably, these respondents would answer the questions the same, regardless of who interviewed them. However, one exception would be if the evaluators were from the host countries, instead of American evaluators. The national identity of the evaluators may be important, given that the program participants (most of them, at least) are aware that the programs are funded by the U.S. government. Additionally, a few of the measures, like those that assess “understanding of the United States” are fairly subjective and difficult to quantify. Therefore, students might answer differently on some of these questions if
they were asked again. Furthermore, if the same questions were asked of the same students and teachers, but at a later date (further from the end of program,) it is possible that the responses would differ.

Similarly, the main problem regarding the reliability of the IVP evaluation is also the timing of the evaluation. If the program participants were asked the same questions a few months or years after their participation, their answers would likely differ. The problem with administering the survey immediately after the program is that the participants are likely to over-emphasize the impacts of the program; they are very focused on the program itself and may not recognize how their other experiences have also influenced their understanding of the United States or their professional goals. Some of the IVP measures are also somewhat subjective. Someone who answered that the IVP increased his understanding of U.S. foreign policy “to a great extent” (a 5 on a 1-5 scale), might answer this question with a “3” or “4” if he were asked again. Additionally, one person may perceive the meaning of “a great extent” differently than someone else. This subjectivity is an issue with all three ECA program evaluations.

The Alhurra performance measures are generally more reliable because they are more clearly defined. For the content analysis, the evaluators developed a code book containing decision rules, to be used during the coding process. Native Arabic-speaking coders were trained according to the decision rules (USC Center on Public Diplomacy 13). The evaluators identified 23 major issues in the news agenda and provided examples of each, clarifying how news stories should be categorized (USC Center on Public Diplomacy 63-64). The coders were also asked whether the programming was value neutral, whether journalists inserted personal judgment, whether opposing views were represented, and whether there was unsubstantiated information.
Each of these terms ("value neutral," "personal judgment," "opposing views," "unsubstantiated information") were clearly defined for the coders.

Inter-coder reliability scores were calculated for each research question and the values were all very high (within the preferred range), with the exception of one score, which was still high enough to be acceptable (USC Center on Public Diplomacy 65-66). Additionally, for the expert discussion groups, standard protocol was developed so that all focus group leaders asked the same questions of participants. The focus group leaders also assured the participants that they would not be quoted by name, in order to generate as open and honest of a discussion as possible (USC Center on Public Diplomacy 11-12, 67-69).

B) The degree to which the evaluation is statistically sound

a) Are the sample sizes big enough to draw relevant conclusions?

It is important to consider both the sample sizes used in these evaluations and the larger population these samples are supposedly representing. While the sample sizes used in the evaluations vary significantly between programs and even within the same evaluation, there are some problems with sample size for all three of the ECA programs, particularly for the samples of teachers and administrators in the Access and BRIDGE evaluations.

For example, the sample of participants used in the Access evaluation were from 2006 program participants in Morocco, Lebanon, Oman, India, Bangladesh, and Pakistan. Since the inception of the Access program, approximately 20,000 students have enrolled in the program in 44 countries throughout the world (Aguirre Division Evaluation of Access Program ix). However, this evaluation is not necessarily trying to make claims about all Access participants; but rather, it represents Access participants, in 2006, from the aforementioned six countries. If this is, in fact, the population that the evaluators seek to represent, then the sample sizes, at least
the aggregate samples for all six countries, are probably not problematic. In total, 613 students, 137 peers, 136 parents, 35 teachers, and 21 administrators were surveyed.

However, when the evaluators look at each region separately, and they look at samples of 4-10 teachers or administrators, these samples do not necessarily produce valid conclusions about the programs. On the other hand, there were not that many teachers to begin with (the population is small), so the samples could not actually have been much bigger (Aguirre Division Evaluation of Access Program xii). The evaluators do not seem to warn against using these small samples for quantitative analysis; however, most of the quantitative data that they discuss comes from students and peers. The only numeric data they present that comes from the teachers’ responses are “Access classroom time allocation by skill or topic,” where they show the average number of hours for the sample of teachers (Aguirre Division Evaluation of Access Program 26, 31, 37, 45, 53). The sample sizes of students in each region are probably big enough for the evaluators to draw valid conclusions about the program. However, we do not know what proportion of the population (all program participants in these six countries in 2006) the 613 students in the sample represent. The evaluators give a range for the number of students in an Access classroom (which differs by country, and can be a rather broad range) and the number of Access sites in a given country, but we do not know the exact total number of students in each country (Aguirre Division Evaluation of Access Program 16).

The evaluators also conducted site visits, interviews, and focus groups. However, these other methods are mainly used to collect qualitative information (quotes from participants), and are used to supplement the quantitative information that comes from the surveys. Therefore, while we do not know what the sample sizes are for the interviews or focus groups, this omission
is not that important. The qualitative results tend to support the quantitative results generated from the surveys, a factor which adds to the overall strength of the evaluation.

For the BRIDGE evaluation, it is also important to define the population that the sample supposedly represents (something that the evaluators do not do very clearly). BRIDGE programs are present in nine different countries in the Middle East, North Africa, and South Asia, yet only four of these countries (Lebanon, Jordan, Morocco, and Pakistan) are represented in this evaluation (Aguirre Division Evaluation of the SCP and BRIDGE Program 6-8). The BRIDGE programs differ between countries. Some countries seem to have more “active programs;” they have teachers who enroll in and complete online courses and students who tend to participate in more regional conferences, physical exchanges, and online forums. The evaluators mention that three out of the four countries included in the evaluation (Lebanon, Pakistan, and Morocco) appear to have some of the most “active” programs (Aguirre Division Evaluation of the SCP and BRIDGE Program 58). Therefore, the programs described in this evaluation may not necessarily be representative of all BRIDGE programs. The relevant population about which the evaluators would be able to draw conclusions, regardless of sample size, would only be program participants in these four countries. Furthermore, the qualitative information gathered from the on-site visit in Lebanon can also not necessarily be generalized to programs in other countries.

A total of 101 internet surveys of students were administered. This sample size appears to be appropriate; however, as with the Access program, it is unclear how many total participants there are in the BRIDGE programs in these countries. The evaluators state that 1,097 students registered on the survey link and began the survey and that 762 students completed 50% or more. Of these 762 SCP and BRIDGE students, 101 were BRIDGE participants. Therefore, we do not know how many of the 1,097 initial respondents were BRIDGE students. Furthermore, this
number also does not indicate the total number of program participants, since not all participants even started the survey (Aguirre Division Evaluation of the SCP and BRIDGE Program 14). There may also be response bias; those who responded to the survey may be the participants who had the strongest feelings about the program and were therefore compelled to participate. Another limitation of this sample is that 81 of the 101 respondents came from the program in Pakistan, so it is unclear that this sample is a good representation of students in the four different countries. At the same time, the Pakistani Program may have been much larger than the other programs, explaining why the vast majority of respondents are from that program (Aguirre Division Evaluation of the SCP and BRIDGE Program 111). Furthermore, while the evaluators indicate that there were four countries represented by the surveys, no Jordanian students responded to the survey and only one teacher and one master trainer from Jordan are represented (Aguirre Division Evaluation of the SCP and BRIDGE Program 17).

Thirty-five teachers and 5 master trainers were also surveyed. The sample of teachers may be large enough to draw relevant conclusions, but again, like with the student sample, the majority of teachers are from Pakistan. The evaluators do warn, however, that “the sample size for the BRIDGE teachers is very small and percentages should be interpreted with caution. Results are illustrative only” (Aguirre Division Evaluation of the SCP and BRIDGE Program 133). They further caution that the small sample size “is not representative of the entire population of BRIDGE teachers” (Aguirre Division Evaluation of the SCP and BRIDGE Program 157). Because of the small number of master trainers, they were grouped with master trainers from other SCP’s.

The issue of sample size in the IVP evaluation is more straightforward. The evaluation represents the population of 95 participants from 19 different countries in the IVP Special
Initiative projects in 2002. The sample includes 85 respondents, 89% of the total population (Office of Policy and Evaluation 4). The questions that are program-specific have small sample sizes, which makes sense given that the biggest program had 20 participants. While many of these samples are small, some of them include 100% of the participants in that program. Therefore, it may not be problematic to draw conclusions about these programs from these small samples. However, it is unclear how useful it is to simply summarize how program participants felt about the program, without using the data to draw broader conclusions about the potential of the program in the future. It would make sense that the evaluators would want to use these statistics to somehow demonstrate how leaders from these regions would generally respond to such programs. Drawing such general conclusions from this data is probably not valid, and while the evaluators do not explicitly make such conclusions, they do hint at the broader implications of their research.

The issue of sample size for the Alhurra evaluation is distinct. The bulk of the evaluation is based on the quantitative content analysis of 4,622 minutes (77 hours) of Alhurra news programming from November 2007. This is a rather large sample size; however the programming from November 2007 may not necessarily be representative of Alhurra programming year-round (USC Center on Public Diplomacy 13). There may have been particular political, economic, or social events or conditions during that time that influenced the content and objectivity of the programming. The balance of issues and opinions presented may be different during other months.

The qualitative part of the analysis came from expert discussion groups in Beirut, Cairo, and Dubai, which had 12, 19, and 11 participants respectively. The evaluators used the
discussions from the focus groups to enhance their quantitative analysis; since they did not attempt to analyze these results numerically, sample size is not an issue.

b) Is the phrasing of the questions likely to elicit biased responses?

The possibility of biased responses due to question phrasing is a slight problem in all four evaluations. As I discussed in the validity section earlier in the empirical analysis, the majority of questions are phrased appropriately. However, it is important to note that many participants were asked the survey questions in another language and the English translations of the questions that appear in the evaluations may have different connotations than those in the original language. Furthermore, for all three ECA programs, respondents may have felt pressure to respond positively to questions about improvement in English skills or changes in understanding of the United States after program participation. Presumably, Access and BRIDGE participants assume that their English should improve significantly after participating in these programs, given the time they have invested. They might feel pressured to over-emphasize the success of the program and they may have exaggerated their improvement to appear smart and hardworking for the American evaluators. When IVP participants were asked, “how do you intend to use this new information or experience when you return home?,” they likely felt pressured to produce a specific answer, because it seemed like this was expected of them (Office of Policy and Evaluation Appendix B 2).

c) Non-random selection of participants

Because participants in these programs were not randomly-selected, one must proceed with caution when using these samples to draw conclusions about program efficacy. Program

---

2 An important distinction should be made here. I am not looking at whether the samples for evaluations were randomly-selected out of all participants in the program (in some cases they were and in others, they were not). The samples used in the evaluation are important in determining the representativeness of the evaluation of all program participants. However, what I am focusing on in this section is that the program participants are not randomly-
participants across all four programs may be categorically different from other members of the target populations. For example, Access students are selected based on written assessment, an entrance test, and an interview. Students have to show initiative and to possess certain skills; they may generally be better students than their peers who also live in underserved areas, who do not participate (Aguirre Division Evaluation of Access Program x). For example, in Morocco, qualified students were recommended by teachers or school administrators and were required to have demonstrated commitment to their studies through achievement in school. Therefore, differences between Access students and their peers are not necessarily due to their participation in the program, but could also be due to other differences between the two groups.

The same is generally true for BRIDGE and IVP. For instance, participants in IVP are leaders in their countries and perhaps have a more nuanced understanding of Americans than the “average” citizen. Therefore, their high level of knowledge about Americans cannot be attributed solely to their participation in IVP. Clearly, by choosing to participate in IVP, the participants are showing their commitment to interacting with professional counterparts in the United States and with other Americans. Their participation in the program may have increased their desire to maintain relationships with Americans but most likely, this desire already existed.

Likewise, the Alhurra discussion groups included opinion leaders who have education or work experience in the United States, who may be different from the “average” Alhurra viewer. While these people may be among “Alhurra’s desired audience,” because of their influence in their communities, they do not necessarily represent all Alhurra viewers (USC Center on Public Diplomacy 38). Furthermore, because they chose to participate in the focus group, they may be more likely to have critical views of Alhurra programming than other Alhurra viewers.

---

selected; they differ in important ways from other members of the “target population.” Therefore, differences between program participants and non-participants are not necessarily due to participation in the program. These groups may have just been different to begin with.
C) The strength of causal inference and the interpretation of the data

a) Do the evaluators consider questions of causality, bias, confounding factors and mediator/moderator variables?

Issues involving causal inference and confounding factors have already been hinted at in other sections of this analysis (for instance, in the section on non-random selection of participants, where it was suggested that participants and non-participants could differ in many ways). As has already been discussed, the evaluators, at least for the three ECA programs, sometimes overstate the positive impacts of the programs.

For example, in the summary of evaluation findings for the Access program, the evaluators state that “more than 80% of Access students believe they have developed good or excellent leadership skills due to the Program” (Aguirre Division Evaluation of Access Program xi). The first problem with this statement relates to the validity of the performance measure; this question may measure the perceptions of Access students regarding their leadership skills, but it does not measure their actual leadership skills. However, another problem with this question relates to the invalid assumption of a causal relationship between the program and students’ leadership skills. Upon further analysis of the evaluation, it is clear that the data upon which conclusions are drawn about leadership skills are comparisons of measures of perceived leadership skills between Access students and peers, not changes in perceived leadership skills of Access students due to their participation in the program. While it is true that more than 80% of Access students rate their leadership skills as excellent or good, similar percentages of peers also rate their leadership skills as excellent or good. In some cases, the Access students rated themselves slightly better than their peers, but this does not mean that their participation in the Access program caused that difference (or that the difference is in fact real and not just perceived).
Therefore, it is unclear how the evaluators have concluded that these highly-rated leadership skills are “due to the program” (Aguirre Division Evaluation of Access Program 61-62, 80-81). The higher motivation or ambition of Access students could possibly be a confounding or mediator variable.

While the problem of improperly implying causality is still somewhat of an issue in the BRIDGE evaluation, the BRIDGE evaluators seem to do a better job addressing the concerns about confounding factors. For example, the evaluators note that the BRIDGE students report more change than the SCP students in their understanding of the United States as a result of the program. They acknowledge that, although this difference could be due to differences between the programs, it could also be due to “something else” like “cultural norms, levels of knowledge base in these countries, pre-existing access to information about the United States in these countries, cultural propensity to report positive responses,” and not the “program structure itself” (Aguirre Division Evaluation of the SCP and BRIDGE Program 116).

In the IVP evaluation, as well as in the other two ECA evaluations, many of the questions are phrased using the basic format, “as a result of the program…how did x change?” For example, the IVP evaluators ask respondents, “to what extent did your IV Program experience help to increase your understanding of the following aspects of the United States?- U.S. government and politics, U.S. values, U.S. foreign policy, diversity in the United States etc. (Office of Policy and Evaluation Appendix 8). This is generally a good structure for questions. Nonetheless, just because the question is asked this way does not necessarily mean that the participants’ knowledge, skills, or understanding are due solely to the program. As was previously mentioned, without doing a) a pre and post-program evaluation to measures changes in participants’ skills or knowledge levels or b) a comparison between participants and a non-
participant control group (a group somehow similar to the participant group in all ways except for their participation in the program), it is difficult to show that the program actually caused a specific change. Other characteristics of IVP participants, such as their high education levels, may contribute to their open-mindedness and positive views about the United States. The IVP evaluators do not acknowledge these confounding variables. Instead, they claim that the evaluation results “clearly demonstrate” that the participants “experienced many profound changes in their views and perceptions of the United States and Americans as a result of their IV experience” (Office of Policy and Evaluation 16). I do not doubt that this program could have that type of profound impact on participants; I simply question whether this conclusion has been proven by the evaluation. The evaluators should have provided the caveat that the views and perceptions of participants are not only the product of their participation in the program, but are also influenced by other factors.

Unlike the other three evaluations, the Alhurra evaluation is less about showing causality and more about coding and analyzing different aspects of Alhurra broadcasting. At the same time, through their analysis, the evaluators are still trying to demonstrate whether or not Alhurra is accomplishing its goals. There are certain trends that the evaluators could perhaps explore in more depth. For example, the evaluators imply that the fact that Alhurra was more likely to promote U.S. policy than describe it neutrally is somehow in conflict with its goal to provide clear and effective presentation of the policies of the United States. Furthermore, they suggest that, “by relying on U.S. or Western sources for western-related issues and interests and Arab or other sources for Arab issues, Alhurra may be seen as just another old-style channel that much of the Arab audience now considers obsolete” (USC Center on Public Diplomacy 32).
Another concern of the evaluators is that Alhurra’s coverage was more likely to be positive when the issue at hand was directly tied to current U.S. policy goals in the region, a trend that causes audiences to see Alhurra as propagandistic and less credible (USC Center on Public Diplomacy 36). For instance, in reporting about Arab-Israeli relations and the Iraq War, pro-Western and anti-Arab expression was much more likely than vice versa. The evaluators then point out that “any news organization that seeks to appeal to an Arab audience but is perceived to be pro-Israeli and an apologist for the American military presence in Iraq will suffer credibility problems” (USC Center on Public Diplomacy 24). While these concerns are certainly valid, they do not necessarily prove that Alhurra is failing to accomplish the program goals, as the evaluators imply. It is true that Alhurra strives to be an objective source and to be attractive to viewers, but it also strives to promote American policies. While Alhurra should be objective, broadcasts should also represent U.S. policies and, the fact is, the United States is generally supportive of Israel, a policy that reporters should not necessarily deny. While the BBG certainly does not want to lose viewers due to lack of credibility, they also do not want to mislead viewers about our policies simply to curry favor with them.

b) **Was there any sort of quasi-experiment (comparison across multiple conditions) or pre and post test?**

The only evaluation that contained any sort of quasi-experiment was the Access evaluation, with the inclusion of a “peer group” for some of the performance measures. The peers were mainly friends of Access program students or other youth in the community who were available for interviews. Peers were selected through the interviews with the Access participants. Each participant was asked to provide the name of one peer whom the evaluation team could interview (except in Oman and Bangladesh), lists of peers were compiled for each
city, and a random sample was selected. In Oman, peers were selected by Access administrators and in Bangladesh, no peer interviews were conducted at the request of the U.S. embassy (Aguirre Division Evaluation of Access Program 7). Presumably, these peers did have a lot in common with Access participants. The evaluators do a relatively good job choosing the peer group, since they are from the same community as the participants and have similar amounts of schooling (Aguirre Division Evaluation of Access Program 13-14). Nonetheless, the peer group is not a perfect control group and the evaluators still do not know whether there are general differences between the two groups.

Furthermore, none of the evaluations show any pre-program analysis. The Access participants were asked to assess their English skills before the program, but they assess those skills after already having participated in the Access program (so they did not actually fill out a survey before the program started) (Aguirre Division Evaluation of Access Program 63). The IVP evaluators state at the beginning of their report that they “developed the survey questionnaire and pre-tested it with participants of one project,” but then they never mention results from this pre-test anywhere in the evaluation and do not use those results when drawing conclusions about the program (Office of Policy and Evaluation 5).

c) Do the evaluators compare this program to other programs using cost-benefit or feasibility analysis?

The only evaluations that reference other programs are the BRIDGE and Alhurra evaluations. The BRIDGE evaluation is actually an evaluation of two distinct programs, BRIDGE and SCP. The evaluators assess the two programs using essentially all of the same performance measures. They do compare results from the two programs and point out when the results differ significantly. For example, BRIDGE students believe they have less basic
knowledge of U.S. values and culture than SCP students, and BRIDGE students report more
change than SCP students in their understanding of the United States as a result of the program
(Aguirre Division Evaluation of the SCP and BRIDGE Program 115-16). However, the
evaluators do not compare BRIDGE to SCP using a cost-benefit or feasibility analysis.

The Alhurra evaluators often compare Alhurra to Al Jazeera, comparing news content
and on-the-ground resources of the two organizations. They also discuss viewers’ perceptions
that Al Jazeera is more credible and objective. The evaluators draw significantly upon research
on Al Jazeera in their analysis (USC Center on Public Diplomacy 17-18, 32, 35, 42, 45, 46, 48-
49). The evaluators compare viewer assessments of the two television programs but they do not
do an in-depth cost-benefit or feasibility analysis.

D) The relative neutrality of the evaluators: Who are the evaluators and what stake do
they have in the evaluation process?

None of the evaluators appear to be overtly biased. However, it is nevertheless important
to consider their unique perspectives and incentives when they are evaluating these programs.
The Access and BRIDGE programs are both evaluated by the Aguirre Division of JBS
International Inc., a private contract evaluator hired by the government. In both cases, research
assistance was provided by in-country partners (research organizations based in the program
countries) (Aguirre Division Evaluation of the SCP and BRIDGE Program iii; Aguirre Division
Evaluation of Access Program v). These research partners provided local expertise about the
communities studied, translated and piloted survey protocols in local languages, and provided
Aguirre with feedback on the questionnaire pilots. Protocols were then revised accordingly. The
in-country partners coded and translated all of the data and Aguirre then did the data analysis
(Aguirre Division Evaluation of Access Program 6).
Presumably, because JBS International is an independent research company, the employees are objective evaluators, independent from government influence. However, JBS contracts are funded by the government and the recipients of these contracts are probably hoping to be re-hired. While the evaluators do not necessarily appear overtly biased, it is important to consider their position, and their desire to ensure additional contracts. One must also distinguish between the paid evaluator, and the person who negotiates a contract. The people actually evaluating the program may want to state one thing, but the person responsible for selling the business might curtail certain comments. That being said, JBS appears to be a very professional organization with a lot of experience researching and evaluating government programs and has worked for a wide range of government, business, and nonprofit clients. Therefore, it is unlikely that the evaluators would bias their evaluation, at least not intentionally ("JBS International: Professional Services").

In contrast, the IVP evaluation was completed by the Office of Policy and Evaluation at ECA (ECA/P). Unlike JBS, ECA/P is an internal evaluation unit in the government and is actually a part of the same bureau that is responsible for the program. As was mentioned in the methodology, there may be reason to question the legitimacy and credibility of an evaluation that is funded from the same agency budget which funds the evaluated program. Especially at a time when public diplomacy programs have come under scrutiny for not having proper evaluation techniques in place, ECA/P may feel pressured to publish results that prove the efficacy of the program. The ECA/P may only publicize the positive results of the evaluation, in an attempt to sustain or increase funding for the program. The evaluation is overwhelmingly positive and is certainly less detailed and complex than the JBS evaluations. The positive results do not necessarily imply that the evaluators were unreliable, but it is as least possible that they over-
emphasized the success of the program. Furthermore, an internal government organization is more likely than independent evaluators to know the measurement criteria that the OMB uses to determine how much money is given to each bureau and program. Therefore, it is possible that such evaluators would focus their efforts on these criteria and ignore other indicators that could also be important. The IVP evaluation does include questions that cover most aspects of the program, so the exclusion of certain criteria does not appear to be a problem for this evaluation.

The Alhurra evaluation was completed by an independent, academic institution, the USC Center on Public Diplomacy at the Annenberg School. Although independent institutions may often have the privilege and discretion to choose and prioritize the objectives of the evaluation, this does not seem to be the case with the Alhurra evaluation. The evaluators do not evaluate Alhurra purely for academic research; they were contracted by the BBG to determine whether Alhurra broadcasts were conforming to the standards of the U.S. International Broadcasting Act and MBN’s Journalistic Code. The goal of the evaluation was to determine “whether Alhurra TV provides viewers in the Middle East with a reliable source of accurate, objective, and comprehensive news and information, and to determine if Alhurra TV presents United States policies and responsible discussions of those policies” (USC Center on Public Diplomacy 8). The evaluation is very focused on answering the questions posed by the BBG about program efficacy. Therefore, although the evaluator is an academic institution and not a private company, the approach and goals of the evaluators are actually quite similar to those of JBS.

E) The examination of the program in a broader geographical, social, and cultural context

a) Do the evaluators acknowledge how cultural, religious, and social norms of the program setting can impact program success?
The evaluators explore the impact of the program setting on program success to varying degrees. The Access evaluation is the most nuanced, while the BRIDGE and IVP evaluators sometimes ignore the unique characteristics of participants in particular regions or countries and the cultural norms of each program setting. The Alhurra evaluators also fail to consider the diversity of the Alhurra audience.

The results from the Access evaluation are shown by country and by region (Near East and South and Central Asia). The evaluators acknowledge the social and cultural differences of different program settings and state how these differences impact the set-up of the program, the impacts of the program, and the evaluation process. For instance, parents and peers were not interviewed in Bangladesh upon the request of the U.S. embassy; parents were not interviewed in Chennai, India, because the students were residents of a charitable boarding school which serves children without parental care (Aguirre Division Evaluation of Access Program 12-13). Clearly, the lack of data from parents in these programs influences the conclusions that the evaluators are able to make when compared to the other programs, from which parental interviews were analyzed. Furthermore, the Access program in Chennai is likely to have a different impact on students than would programs at “regular” public schools, where students go home to their parents every night. A few schools also had particularly low Access retention rates, due to class schedule conflicts, the difficult commute to classes, and competition with other family responsibilities. Students who do not complete their Access classes will likely have a different experience than those who do (Aguirre Division Evaluation of Access Program 47).

The evaluators also explain the basic structural differences between the programs in the two regions as well as between countries within the same region. They effectively tease out the factors that contribute to differences in the programs and recognize how these differences affect
the impact of the programs on the students. For example, while most NEA (Near East) participants attended their Access program classes outside of their regular school, most of the SCA (South and Central Asia) participants attended Access classes at their regular schools, usually after school hours. The participants were also at varying stages of the program (Aguirre Division Evaluation of Access Program 11-12). Furthermore, the evaluators looked at the gender composition of each program; some were fully integrated (Morocco and Lebanon) and others were gender-segregated by class or within each class (Oman, India, Bangladesh, and Pakistan).

The programs cater to the needs and cultural norms of each country and school. For instance, in Oman, the program directors had to create a more basic level class because the starting level of students’ English was lower than it was in other countries (Aguirre Division Evaluation of Access Program 31). Certain countries also had differing activities because of security concerns or cultural norms. For example, in Lebanon, due to security concerns, there was a limited amount of direct interaction with Americans, in comparison with Morocco, where travel was unrestricted and there were many teachers and visitors from the United States with whom students could interact (Aguirre Division Evaluation of Access Program 27).

The evaluators also mention ways in which the programs could perhaps be more culturally-sensitive. For instance, teachers in the NEA region mentioned that certain topics that U.S. students might consider mundane are taboo in more conservative Muslim populations. One teacher mentioned a textbook that discussed wearing bikinis on the beach and the need to stay in shape in order to look good in a swimsuit. The students were curious about the reference but the teacher felt the need to be cautious and avoid offending the students personal or religious beliefs (Aguirre Division Evaluation of Access Program 61). These are just a few of the many factors that the evaluators looked at when comparing programs between countries.
The BRIDGE evaluation is not divided by country like the Access evaluation and results are reported for all BRIDGE participants (from four different countries) together. This difference is partially logistical; the sample of students in the Access evaluation was much bigger, making it feasible for the evaluators to split up the results by country. As was mentioned earlier, because the majority of students who answered the BRIDGE student survey were from Pakistan, the sample sizes from Lebanon, and Morocco would have been too small to do any sort of quantitative analysis (Aguirre Division Evaluation of the SCP and BRIDGE Program 17).

Although the evaluators acknowledge that the programs vary in different countries, they do not discuss whether, and if so how, the data from Lebanon and Pakistan (where the majority of the data comes from) may differ from data about all program countries. At the end of the evaluation, when discussing their recommendations, the evaluators mention that a “one size” approach does not fit all countries and regions; nevertheless, they argue that certain suggestions are intrinsic to any successful program (Aguirre Division Evaluation of the SCP and BRIDGE Program).

The IV programs are set in various locations throughout the United States. The evaluators list 30 different U.S. cities that the participants visit, but it is unclear how many cities each participant actually visits during his or her three-week program (Office of Policy and Evaluation 2-4). The participants indicate that the program did not show them the full diversity of the United States and that they would have liked to see a greater variety of communities (Office of Policy and Evaluation 19). Based on what the IVP (now called the International Visitor Leadership Program) website says about current programs, the location of the program depends on the program focus and the profession of the participants ("International Visitor Leadership Program"). However, this may be different for the Special Initiatives Projects. Since certain programs are probably focused in certain areas of the country, it would have been interesting to
see if programs in certain locations were more effective than others. Furthermore, each program is made up of participants from various countries, and the participants’ country of origin could affect their program experience (something that the evaluators do not address).

An inherent part of the Alhurra evaluation is to determine whether Alhurra is adequately meeting the needs of its viewers. Nevertheless, the Alhurra viewership is rather diverse; Alhurra broadcasts in 22 different countries and the evaluators do not consider how the programming may be received differently by viewers in different regions. While they do hold focus groups in three different cities, Beirut, Lebanon, Cairo, Egypt and Dubai, United Arab Emirates, the opinions voiced by participants in these focus groups do not necessarily reflect the perspectives of viewers in Morocco, Algeria, Somalia, Saudi Arabia, or Iraq, let alone the opinions of the majority of viewers in the cities where the focus groups were held. The qualitative results from these small focus groups cannot be generalized to a larger population.

b) **Do the evaluators consider how the program differs from others of its kind? Do they generalize results to all similar programs?**

The evaluators do not generalize results to all similar programs, although some of them do make comparisons to other programs. The Access evaluators do not compare the program to others of its kind, nor do they generalize the results of the evaluation to other similar English learning experiences. The BRIDGE program is compared to the SCP throughout the evaluation and the evaluators point out when the results differ for the two programs. The IVP evaluators do generalize their results; they imply that the success of the IVP shows the “value and power of people-to-people exchanges in promoting professional collaboration, mutual understanding and global peace and security” (Office of Policy and Evaluation 16). This quote implies that the evaluators use the IVP evaluation to show the value of all people-to-people exchanges, a claim
they cannot accurately make using the data from one evaluation. This exchange program is different from others, in that the participants are prominent leaders in their countries. The effect of this program is likely to be different than that of other exchange programs.

The Alhurra evaluators do compare Alhurra to other news programming available in the region, particularly to Al Jazeera. While the two programs perhaps have some of the same goals, the major difference between the two is that Alhurra is run by a U.S. organization, the BBG, and has a specific responsibility to discuss and promote U.S. policies. Therefore, the two programs are not entirely comparable because they have different objectives.

c) **Time frame of the evaluation: During, shortly after, or long after program participation?**

All four evaluations were administered during or shortly after program participation. The majority of students who responded to the Access surveys were current students; however, some were alumni who had recently completed the program. The current students could have a different perspective than the alumni, who may have had a chance to pursue other schooling or join the work force, and see the effects of their Access training on their lives (as opposed to just predicting these effects). At the same time, some of these alumni had only finished the program a few months earlier, and it may have still been too early for them to truly gauge the long-term impacts of the program. The evaluators do not indicate exactly how long it had been since the alumni had completed the program, but they do point out that some of the Lebanese alumni had still been students during the on-site evaluation, and had finished the program by the time the surveys were distributed. Furthermore, the programs all began in 2003 or 2004 and the evaluation was conducted in 2006, so the alumni could not have been more than two or three years out of the program. While the evaluators do sometimes break the results into two groups
(alumni vs. current students), the number of alumni is relatively small in comparison to the entire student sample, and is not necessarily representative of alumni responses to the program (Aguirre Division Evaluation of Access Program 9).

The BRIDGE evaluators do not indicate when the evaluation occurred in relation to program completion, but it seems that respondents were surveyed shortly after the program (and on-site visits were obviously during the program). The BRIDGE program began in 2002, the evaluation process began in 2004, and the surveys were most likely administered in 2005. Because the surveys were made available to all users of the school’s internet learning centers, the respondents were likely both recent alumni and current participants, as was true of the Access respondents (Aguirre Division Evaluation of the SCP and BRIDGE Program 13-15).

Most IVP participants completed their survey at the end of their stay in the United States and a few completed it back in their home countries. For the Alhurra evaluation, the coding was completed after the coders viewed the programming and the focus groups discussed programming immediately after viewing certain clips (USC Center on Public Diplomacy 67-69). Although the Alhurra evaluators do not explicitly state the time-frame for the quantitative analysis, it is safe to assume that the coding was completed shortly after the coders viewed the programming, as to assure the most accurate results (USC Center on Public Diplomacy 67-69).

**d) Do evaluators distinguish between short-term and long-term effects of the program?**

Since the three ECA program evaluations are conducted immediately or shortly after program participation, they do not sufficiently capture the long-term impacts of the programs. The data generated from the Alhurra focus groups also do not reflect long-term outcomes. In some instances, the evaluators recognize that the evaluation only measures short-term effects, but other times, particularly in the IVP evaluation, they tend to overlook this problem.
For example, the Access evaluators state upfront that, “this is being referred to as a pilot evaluation because it was both formative and retrospective; it was designed to gather information about what has been done in order to inform future Program management decisions, and does not represent the full impact this Program will have over time” (Aguirre Division Evaluation of Access Program 1). Additionally, the BRIDGE evaluators also addressed long-term impacts by developing strategies for program sustainability after talking with participants during on-site visits and analyzing responses to the online surveys. They explored the following options: establishing and maintaining school networks, establishing and maintaining professional collaborations and relationships, maintaining contacts with people met during the program, collaboration with and support from other donors in program countries (so that the program could continue after U.S. funding ceased), information dissemination (maintaining awareness of program activities, services, and impacts in local communities), and fundraising and planning to meet resource needs (Aguirre Division Evaluation of the SCP and BRIDGE Program 159). The evaluators also tried to capture the more permanent, far-reaching impacts of the program by assessing multiplier effects. They assess how the program affects the community as a whole, not just the program participants. Nevertheless, the data they use to determine these multiplier effects come from the same surveys as all the other data in the evaluation, gathered shortly after program completion.

Throughout the IVP evaluation, the evaluators do not recognize the limitations of the data due to the time-frame of the evaluation. The evaluators indicate that participants have been truly transformed; however, the degree of this transformation cannot really be assessed with this evaluation. Nonetheless, the evaluators do concede towards the end of the report that “the evaluation results are based on immediate assessments…Therefore, it is recommended that a
follow-up study be conducted in the future with the same participants in order to learn if these assessments (i.e. participants’ new understandings and changed perceptions) are sustained over time and to learn whether the International Visitors proceeded with their planned future activities, as reported in this evaluation” (Office of Policy and Evaluation 16)

The problem of distinguishing between short and long-term impacts does not seem to be as much of an issue for the Alhurra evaluation. Because the majority of the analysis was based on coding, it makes sense that this analysis was completed immediately after viewing the programming. The coders were trying to determine the issues that the programming addressed, and this is done most accurately when the programming is fresh in their minds. They were not trying to assess the long-term impacts on the viewers with this quantitative analysis.

Distinguishing between short and long-term impacts, however, could be a problem with the focus group discussions. The focus group participants viewed a clip from Alhurra programming and were then asked a series of questions about the clip immediately after viewing it. The way in which the participants react to the clips only demonstrates the short-term impact of the programming. Although the participants were instructed to answer the questions by drawing not only from the clip shown, but also their prior viewing experiences, they are likely to focus on what they viewed most recently (USC Center on Public Diplomacy 67). The evaluators do not explain how the immediate responses of the participants could differ from the long-term impacts of Alhurra. What viewers ultimately take away from watching Alhurra over time, versus the immediate reactions of focus group participants who were asked to concentrate on one specific clip, could differ significantly.

e)  **Do evaluators look at results from other similar programs to see whether answers are generally reliable?**
In general, the evaluators do not look at program evaluations of similar programs to see whether the results have proven to be reliable. There are other programs for students that are similar to Access and BRIDGE that the evaluators could have looked into, but they do not mention doing such analysis. At the same time, as was highlighted earlier in this paper, the number of available evaluations of public diplomacy programs is limited. Since the IVP has been around for awhile, the evaluators draw upon the results of the overall IVP when discussing the Special Initiatives Projects. For instance, the IVP evaluators acknowledge that many IVP participants in the past have gone on to become very influential, well-known international leaders. In the introduction to the evaluation, the evaluators mention that “ECA’s International Visitor (IV) Program has a long-standing history of strengthening the United States’ international relations by bringing current and future leaders to the United States.” Furthermore, “as of June 2002, 212 former International Visitors had become Heads of Government, including: Hamid Karzai (Afghanistan), Anwar Sadat (Egypt), Tony Blair and Margaret Thatcher (United Kingdom), Gerhard Schroeder (Germany), Kim Dae-Jung (South Korea), Oscar Arias Sanchez (Costa Rica) and Ricardo Lagos (Chile)” (Office of Policy and Evaluation 1).

Because the BBG has programming throughout the world and many of their programs have been around for much longer than Alhurra, the Alhurra evaluators could have conceivably looked at results for other BBG programs. However, they do not mention any such analysis in the evaluation.
V. Conclusion

After critically examining these four evaluations, I have developed a few main conclusions about the evaluation process and how and why these programs have accomplished or failed to accomplish their stated goals. These conclusions can be divided into three over-arching categories. The first category addresses the difficulty of assessing the full impact of each of these programs in just one evaluation. This difficulty can be attributed to the challenge of showing a causal relationship between program participation and outcomes, and the importance of long-term impacts and multiplier effects, which are difficult to capture in evaluations that are administered shortly after program completion. My second major set of conclusions deals with the conflicting goals of public diplomacy. On the one hand, some of the evaluators claim that public diplomacy is supposedly a two-way interaction between the United States and foreign populations, but on the other hand, it is sometimes interpreted as propaganda which the United States uses to better its image abroad, rather than to communicate with and learn about other cultures. Finally, I will elaborate on the differences in outcomes between the three ECA programs and Alhurra and will provide some possible explanations for these divergences.

A) The difficulty of assessing the full impact of these programs
The ideas presented in these program evaluations support the suggestion expounded in the literature review concerning the power of personal interactions between citizens of different countries. Many of the participants in the three ECA programs explain how the interactions they had with U.S. citizens really helped them to better understand the United States and to develop a more positive view of the country. Nevertheless, it remains unclear whether these program evaluations actually *prove* that these exchange programs were the cause of changes in understandings and opinions of the United States. The concerns addressed in the literature review about the difficulty of separating the impact of public diplomacy programs from an array of other factors (political, geographic, historical, educational, economic, ethnographic, diplomatic and humanitarian) that also influence public opinion certainly appear legitimate (ExpectMore.gov "Detailed Information on the Broadcasting in Arabic Assessment"). Simply asking a program participant to assess how the program changed his or her perceptions or skills does not necessarily demonstrate the impact solely of the program on these perceptions or skills. Therefore, as was recommended throughout the empirical section, a pre and post-program test could be helpful in all of the ECA evaluations.

The general inability of these evaluations to assess long-term effects is also an issue. All of these evaluations state the need for follow-up reports that can determine whether the program impacts found in the initial evaluation are actually sustainable. The BRIDGE evaluation does attempt to address the issue of sustainability, but it is difficult to fully address the issue without actually going back to participants a few years after program completion and asking them how the program continues to impact their lives. While evaluators recommend follow-up analysis, they do not necessarily state that they plan to perform this analysis. The Evaluation and Measurement Unit (EMU) at the Office of Policy, Planning, and Resources for Public
Diplomacy and Public Affairs (R/PPR) supposedly already designs and implements program evaluations, “to analyze the long term impact of public diplomacy programs that inform, engage and influence foreign audiences” ("Evaluation and Measurement Unit: Office of Policy, Planning and Resources for Public Diplomacy and Public Affairs (R/PPR)"). It would be interesting to see whether the EMU has already completed a number of evaluations and if so, how these evaluations compare to the four assessed in this paper.

Furthermore, the multiplier effects of the program are also extremely important. Many of the evaluations do attempt to quantify the effects of the programs on non-participants, like peers, parents, and the community at large. The Access evaluation does the best job assessing multiplier effects out of the four evaluations detailed in this paper, by including parents and peers in the survey. The BRIDGE evaluation addresses multiplier effects by asking teachers and students the extent to which they shared information about the program with their peers. However, it is more effective to ask those peers directly, as was done in the Access evaluation.

Nevertheless, the true impact of such multiplier effects can only be seen in the long run. For instance, does the number of program participants continue to increase year after year? Is the program well-known and highly regarded in the community? Has the program generally fostered positive attitudes towards the United States, and a desire to learn about other cultures and to participate in exchange programs? Particularly when the program is small and only available to a small number of participants (like IVP), multiplier effects are very important in justifying the impact of the program. One could argue that programs like Alhurra, which are available to a much wider, larger, and broader population than IVP, are more cost-effective in accomplishing the goals of public diplomacy. However, if IVP has a dramatic impact on the participants, who later become leaders who influence entire communities or countries, while
Alhurra only has a moderate impact (and not always a positive one) on viewers, this claim about the superiority of Alhurra might not be true.

**B) Public diplomacy: A two-way process or just American propaganda?**

In 2006, Mohan Dutta-Bergman discussed how American public diplomacy efforts for the last fifty years have largely been one-way communication tactics that the U.S. government uses to influence and change the opinion of foreign publics, while maintaining its unwillingness to change U.S. opinions or policies. Dutta-Bergman critiqued this one-way method, arguing that it generates further misunderstanding and resentment among foreign populations. He defended a more “culture-centered approach” which focused on the relationships between cultures and established an ongoing dialogue centered on respect and the desire to develop mutual understanding (Dutta-Bergman 119). Based on the evaluations of these four public diplomacy programs, although public diplomacy efforts have become more culturally-sensitive over the years, it seems that U.S. public diplomacy may be still using the “sender” (the United States) and “receiver” (foreign populations) model that it has used for decades. After all, the accepted definition of public diplomacy, “the promotion of America’s interests, culture and policies by informing and influencing foreign populations,” is decidedly one-sided (Epstein and Mages).

However, some of the ECA evaluations imply that U.S. public diplomacy efforts are *supposed* to be two-way dialogues. For example, as is discussed in the IVP evaluation, the “ultimate goal of any exchange opportunity is that ‘both sides’ benefit from the exchange of new information and sharing of diverse perspectives and ways of doing things” (Office of Policy and Evaluation 11). The evaluators go on to explain that this logic is fundamental to the design of the Special Initiatives projects and any ultimate success; it is just as important for Americans to learn from and about project participants as it is for the IV participants to learn from and about
Americans. However, the evaluation content does not reflect this argument because the American participants are not represented in the evaluation. The evaluators determine program success largely based on the ways in which the program improved the foreign participants’ understanding and opinion of the United States and of Americans. Based on this evaluation structure, one might question whether the ultimate goal of this program is really to foster a two-way exchange or if it is just about improving the image of Americans abroad. Do we only care whether or not foreigners think that Americans understand their culture and respect their ways of life, or do we also care whether Americans actually do understand and respect other cultures? Perhaps, American public diplomacy programs are more effective when foreigners perceive them to be two-way dialogues, even if the American government is really only concerned with being the “sender,” and not the “receiver.”

Furthermore, the participants in these programs seem to recognize this tension and sometimes question whether the U.S. government has a “hidden agenda” when they fund these exchange programs, especially when the evaluators continually emphasize that is important that the participants understand that the programs are funded by the U.S. government (Aguirre Division Evaluation of Access Program 59). This emphasis indicates that it is indeed significant that the United States is portrayed as a charitable, friendly, helpful ally. Of course, this is not necessarily negative; the U.S. government is supporting education in these countries by funding these programs and participants should be encouraged to be grateful for this support. However, if too much emphasis is placed on U.S. government sponsorship, the participants and community members could become suspect and doubt U.S. motives.

This tension is also evident in the Alhurra evaluation. The idea that Alhurra is seen as propagandistic is one of the main critiques that the evaluators have of the program. Focus group
participants felt that “Alhurra’s reporting, when stacked against competitors in the region, represented a false or tilted perspective of events…While the U.S. policy and viewpoints were often clearly identified, participants thought that they were unpersuasive or included too little explanation.” Furthermore, while Alhurra’s coverage might be “more positive with regard to the possibilities of peace and stability in the region, these attributes were more often seen as evidence of an agenda rather than coverage that provoked a different point of view” (USC Center on Public Diplomacy 5-6).

On the one hand, viewers see Alhurra as promoting U.S. policies, without attention to other viewpoints. However, the idea the Alhurra promotes U.S. policies in and of itself is not necessarily problematic. After all, the promotion of U.S. policies is a fundamental goal of U.S. public diplomacy programs and Alhurra’s purpose really does seem to be one-way, unlike that of the exchange programs. However, the critique of Alhurra seems to be the following: the programming seems to be subjective because reporters support U.S. policies without giving a thorough explanation of what those policies are and why they are superior to other alternatives. Viewers might not critique the objectivity of Alhurra reporters who constantly support American policies if the reporters showed a compelling reason why these policies produce optimal outcomes. However, it is problematic if Alhurra is blindly supporting all U.S. foreign policy without proper explanation and justification. This kind of unconditional support will only provoke resentment and negative feelings among foreign audiences (as it has already done, according to the Alhurra evaluation).

C) Program success: The three ECA programs vs. Alhurra

The three ECA evaluations are all overwhelmingly positive, while the Alhurra evaluation is quite critical. This difference raises the question, are the ECA programs truly that much more
successful or is the difference partially due to the evaluation processes? Perhaps, because the Alhurra evaluation was performed by an academic institution rather than another type of outside contractor or an evaluation unit within the State Department, it was therefore more rigorous and ultimately more critical. The Alhurra evaluation does not necessarily seem to be more in depth than the other evaluations, but the evaluators might approach the evaluation process differently. Also, State Department officials might have strategically chosen to include these three positive evaluations on the State Department web site, while not making other, more negative evaluations publically-accessible. The Alhurra evaluation, on the other hand, was simply posted as a PDF document and was not linked to the BBG website. This perhaps indicates that the Alhurra evaluators had more independence and control over the evaluation process and by extension, that the BBG had less influence on the process that the State Department did.

Alternatively, it might be true that Alhurra is simply not as successful; after all, the ECA programs are much smaller and more focused. An exchange program is likely to have a bigger impact on participants than a news station will on viewers. That being said, if the ECA programs and Alhurra have differing purposes and goals, they could be equally successful, even if the types of impacts they have differ. However, the evaluations indicate otherwise, suggesting that there is something more to the varying results than the differing sizes and focuses of the programs.

While Alhurra, like the other programs, aims to promote U.S. policies and increase understanding among foreign audiences about the United States, it also has a unique goal, to serve as a reliable news source and to adhere to standard journalistic practices. The conflicting goals of Alhurra, to serve as an objective, reliable news source, while also promoting U.S. policies, seem to be incompatible. At least in the eyes of some Alhurra viewers, Alhurra
reporters are seen to be biased and unprofessional if they consistently try to sway viewers towards pro-American viewpoints. The ECA programs are not charged with serving as unbiased, reliable sources of information about the United States. Therefore, it may not be problematic if these programs are “propagandistic,” while this is deemed a major problem for Alhurra. Alternatively, Alhurra may just be out of touch with the needs of the target audience, something that is perhaps less of a problem for the State Department programs. The discussion group participants do indicate that Alhurra too often relied on “official sources” about issues important to the general Arab public, instead of portraying the “voice of the average Arab.” Furthermore, “the paucity of coverage of Islam and Islamic-related issues indicates insensitivity to one of the fundamental elements of most Arabs’ lives” (USC Center on Public Diplomacy 6).

D) A call for improved evaluation techniques for public diplomacy programs

In conclusion, while public diplomacy institutions are beginning to evaluate their programs and determine whether or not they are accomplishing their stated goals, better and more evaluation is still needed. As my research has shown, even those programs that have been comprehensively evaluated have not always been assessed according to valid and reliable evaluation techniques. It is not sufficient to state that public diplomacy programs are important because they help to combat anti-American sentiment and misunderstanding about the United States. Without proper evaluation techniques, it is impossible to truly understand the effect of these public diplomacy programs on both program participants and the larger communities where the programs are located. If public diplomacy institutions want to maximize the positive impact of their programs, they must employ rigorous evaluation methods that isolate the effects of specific programs, to learn which methods and messages are effective and which are not.
VI. Works Cited:

(Note: The first two are the same document and the third and fourth citations are the same document. One is just an abbreviated citation.)


Ford, Jess T., Director International Affairs and Trade. "U.S. Public Diplomacy: Strategic Planning Efforts Have Improved, but Agencies Face Significant Implementation Challenges. Testimony Before the Subcommittee on International Organizations, Human


<http://exchanges.state.gov/ivlp/ivlp.html>.

<http://www.jbsinternational.com/site/Pages/services-eval.aspx>.


Midura, Christopher; Acting Director, Office of Policy, Planning and Resources. "Testimony Before the Senate Homeland Security and Governmental Affairs Subcommittee on Oversight of Government Management, the Federal Workforce, and the District of Columbia." 2008.


<http://draft.eca.state.gov/PDEO/descriptions.htm>.


