



CHARITY NAVIGATOR

SPEA CAPSTONE GROUP
CHARITY NAVIGATOR 3.0

INDIANA UNIVERSITY, SPRING 2012

SPEA V600 CAPSTONE GROUP

SPRING 2012

Executive Group

Laurie Burns
Lindi Lara
Molly McPherson
May Nguyen
Jason Simons
Sarah Townsend

Logic, Results, and Measures

Jennifer Bellville
Adrianne DiTommaso
Ran Duan
Frances Duberstein
Rachel James
Bekir Keles
Lindi Lara

Independent Evaluations and Standards & Certification Mechanisms

Laurie Burns
Amy Countryman
Kristen Dmytryk
Emrullah Durmus
Sarah Greenberg
May Nguyen

Constituent Voice

Seongje Cho
Jungyun Kim
Bethany Lister
Molly McPherson
Huy Nguyen
Elizabeth Walker
Whitney Williamson

Information Intermediary Research

Kristen Dmytryk
Rachel James
Kathryn Meier
Huy Nguyen
Sarah Townsend

Donor Survey

Adrianne DiTommaso
Sarah Greenberg
Jason Simons
Elizabeth Walker

Americorps

Jennifer Bellville
Laurie Burns
Bethany Lister
Kathryn Meier

Faculty

Leslie Lenkowsky
Joseph Palus

Table of Contents

Executive Summary	1
Chapter One: Problem Statement	5
Chapter Two: Methodology	9
Chapter Three: Intermediaries	19
Chapter Four: Element One	39
Chapter Five: Element Two	75
Chapter Six: Element Three	95
Chapter Seven: Element Four	127
Chapter Eight: Assessment of the Charity	161
Navigator Research Tool	
Chapter Nine: Utility Survey	178
Chapter Ten: AmeriCorps Case Study	192
Chapter Eleven: Conclusions	206
Works Cited	
Appendices	

Executive Summary

Charity Navigator, a highly regarded information intermediary, is testing its next iteration of its ratings tool. The new ratings system, called CN 3.0, assesses charities on four elements and uses them to assign each charity an overall score. The four elements broadly assess each charity's commitment to results reporting and the alignment of resources with program operations. The elements are as follows: "Logic, Results, and Measures," "Independent Evaluations, and Standards, and Certification Mechanisms," "Constituent Voice," and "Alignment of Mission and Resources." Each of these elements contains a series of tests to determine whether charities are working to assess and publish results to increase public understanding of the overall impact of an organization. The Indiana University SPEA Capstone Group was tasked with assessing the ratings tool in its current form and providing recommendations to make it more helpful in encouraging charities to implement results reporting measures in the future.

This assessment of the CN 3.0 ratings tool involved six major tasks: researching the current information intermediary field to assess what concepts are being utilized to evaluate charities' results-reporting efforts, applying the CN 3.0 ratings tool to evaluate 90 total charities, utilizing a supplemental set of indicators (Background Research Tool) to obtain additional information about each charity's results-reporting efforts, conducting a phone survey to gain additional context about charities' results-reporting efforts that were not publicly available on their websites, implementing a survey of 536 CN donors to assess the utility of the

ratings tool to donors who would be using it to make funding decisions, and applying the CN 3.0 framework to a case study of the program AmeriCorps Impacting Health Throughout Indiana.

In the data analysis, the Capstone Group evaluates charity trends regarding each element of CN 3.0. Overall, it found that charities are most involved with reporting information related to results logic and measures. Few organizations are publishing independent evaluations or are involved with standards/certification mechanisms that are outcome-based. Very few organizations are collecting and publishing rigorously-collected constituent voice feedback data. The degree to which an organization's mission is aligned with the way it uses its resources is often difficult to determine from published information.

The Capstone Group concludes that CN 3.0 is a useful tool, encouraging charities to further implement results-reporting measures, but that a number of improvements can make the tool more useful as a whole. General concerns with the tool as it currently exists include: the subjective nature of many of the questions, which lead to varied conclusions from different students reviewing the same charity; instructions are unclear and do not lead to a standard research process; the terminology is not easily understood due to the technical jargon or vague phrasing; the intent of the questions is unclear; and the tests' structure is too restrictive.

The Capstone Group recommends taking steps to reduce subjectivity for charities' ratings on each of the elements, provide explicit instructions and a framework for ensuring understanding among raters, clarify the language and terminology in each of the elements, and utilize a ratings scale in the place of binary questions in order to provide more precise indications of charities' results-reporting efforts.

It also found that charities are engaging in some type of results reporting, but that they have concerns with the CN 3.0 framework. The Capstone Group believes that communicating with charities about the importance and relevance of this framework will be essential in alleviating charity fears and hesitations. Charity Navigator can work to educate and collaborate with charities as they transition into using the tool.

Furthermore, the Capstone Group concludes that CN members are looking for the CN 3.0 concepts when making a decision to donate. While members are seeking this information, they are also confident in turning to Charity Navigator to make a judgment call about the information, rather than looking for the information on the charities' websites themselves. Finally, members do share some concerns about charities' capacity, as well as a limited understanding of the constituent voice concept. Charity Navigator should educate members on the concept and framework of CN 3.0 as it begins utilizing the tool.

When implementing this framework in the case study of AmeriCorps Improving Health Throughout Indiana, the students who did this case study find that CN 3.0 is relevant to the AmeriCorps evaluation program. Though the AmeriCorps program does not currently pass the tests as recommended, there are certain steps the program can accomplish to improve the utility of the tool for its own results-reporting.

This chapter describes in detail case study research methods and findings. It provides detailed recommendations for how each of the elements of the tool can be improved to reduce the challenges for volunteer raters and to ensure that the scores assigned to charities align with their overall level of commitment to results reporting. The students in the Capstone Group who did this case study believe that CN 3.0 can be an effective tool to encourage charities to

implement results-reporting methods to increase their overall impact and ultimately to direct donations to nonprofits that seek to be more effective through results-reporting.

Chapter One

Problem Statement

Evolution of Charity Navigator

Overview: Charity Navigator is highly regarded as one of the premier rating organizations in the entire charitable sector. Its homepage, asserts, “Charity Navigator, America's leading independent charity evaluator, works to advance a more efficient and responsive philanthropic marketplace by evaluating the Financial Health and Accountability and Transparency of America's largest charities” (Charity Navigator, 2012). Charity Navigator’s goals are to “ensure that charitable giving keeps pace with the growing need for charitable programs” (Charity Navigator, 2012).

CN currently rates charities according to two factors:

Charity Navigator 1.0: Charity Navigator 1.0, evaluates nonprofits’ financial health, was the client’s first iteration of its charity evaluation model. This tool used only financial metrics to assess the vitality and sustainability of a charity. The seven metrics were broken into two dimensions: Financial Efficiency Performance and Financial Capacity Performance Metric (Charity Navigator, 2012). They include:

- Financial Efficiency Performance Metrics:
 - Measures of Program Expenses
 - Administration Expenses
 - Fundraising Expenses
 - Fundraising Efficiency
- Financial Capacity Performance Metrics:
 - Primary Revenue Growth

- Program Expenses Growth
- Working Capital Ratio

Charity Navigator relies heavily on Internal Revenue Service Form 990 data for these metrics.

Charity Navigator 2.0: Charity Navigator 2.0 expanded the client's rating model to include accountability and transparency metrics. Seventeen metrics compose this portion of the charities' rating, and include 12 from the Form 990 and five from the reviewed organization's website (Charity Navigator, 2012). These metrics include the following elements:

- Accountability Metrics (from Form 990):
 - Independent Board
 - Material Diversion of Assets
 - Audited financials prepared by independent accountant with an audit oversight committee
 - Loan(s) to or from related parties
 - Documents Board meeting minutes
 - Provided copy of Form 990 to organization's governing body in advance of filing
 - Conflict of interest policy
 - Whistleblower policy
 - Records retention and destruction policy
 - CEO listed with salary
 - Process for determining CEO compensation
 - Compensates Board
- Transparency Metrics (from charity website):
 - Board members listed
 - Key staff listed
 - Audited financials
 - Form 990
 - Privacy Policy

Charity Navigator 3.0 Project Goals and Objectives

In the future, Charity Navigator 3.0 aims to incorporate a third dimension in the rating metrics: results reporting. On the “About Us” section of its website, Charity Navigator states the goal of CN 3.0 is to “completely transform the landscape of charitable giving,” by affecting a “sizable increase in giving to higher performing nonprofits” and “measurable improvement in human welfare and acceleration in solutions to our world’s most persistent problems” (Charity Navigator, 2012).

Charity Navigator aims to utilize four distinct elements to assess charities’ commitment to results reporting:

- Element One: Logic, Results, and Measures
- Element Two: Independent Evaluations, Standards and Certification Mechanisms
- Element Three: Constituent Voice
- Element Four: Alignment of Mission and Resources

These are described further in subsequent chapters of the report.

Progress to Date: Charity Navigator has been actively working on CN 3.0 since 2010 (Berger, 2012). Throughout 2010, the client developed an initial CN 3.0 prototype with funding from Hewlett Foundation. In the fall of that year and the spring of 2011, the initial CN 3.0 prototype, an online research tool, underwent two rounds of student testing which indicated that the initial indicators were too rigorous for the current status of charity results reporting. In the summer months of 2011, Charity Navigator revised the prototype based on student raters’ feedback. It sent out a letter and revised concept note to universities in the fall to request input and student researchers. In the winter and spring of 2012, a group of students from Indiana

University's School of Public and Environmental Affairs engaged in testing and analyzing the second iteration of the CN 3.0 Research Tool.

Indiana University SPEA Capstone Group Role: Throughout this project, the SPEA Capstone Group worked with Charity Navigator and its consultant, Keystone Accountability, to test the CN 3.0 Research Tool. This report details the group's experiences, conclusions, and recommendations for CN 3.0 based on that experience and additional research.

Chapter Two

Methodology

Overview and Capstone Group Structure

Approach: The Capstone Group used a multipronged strategy to conduct the Charity Navigator 3.0 Capstone Project, involving four overarching tactics:

1. Background research: Research on methodology and approaches of other information intermediary organizations, especially focused on their use of concepts similar to CN 3.0's elements
2. Charity data collection: Research on a sample of 90 CN charities' websites using the Charity Navigator 3.0 Research Tool and a Background Research Tool, developed by the Capstone Group. Observations on this process were recorded in student journals
3. Feedback from Charity Navigator constituents: Telephone interviews with CN charities' leaders and online survey of CN members
4. AmeriCorps Improving Health Throughout Indiana case study: Consultation with AmeriCorps program and application of CN 3.0 Elements to 17 AmeriCorps host site programs

Task Division: Students worked together as a group to perform charity data collection during the research phase, which lasted from mid-February through the first week of March 2012. They then divided into six teams of four to five people to address different portions of the analysis and recommendations:

- Team One: Element One (Logic, Results, and Measures)
- Team Two: Element Two (Independent Evaluations, Standards and Certification Mechanisms)
- Team Three: Element Three (Constituent Voice)
- Team Four: Outside Research (Methodology of other intermediaries)
- Team Five: Utility (Charity phone surveys and CN member survey)
- Team Six: AmeriCorps Improving Health Throughout Indiana (Case study in Results Reporting)

Individuals from Teams Four, Five, and Six also served as representatives to Teams One, Two, and Three to create cross-team communication and collaboration.

Outside Research

Research Methods: Team Four students researched practices used by a variety of information intermediary organizations. It compared ratings and standards methodology and current trends across the field, as they relate to CN 3.0's four elements. Students gathered information on 19 groups.

Analysis: Team Four identified a sample of ten organizations whose mission and methods most closely matched Charity Navigator. It analyzed trends and patterns in the sample's work to inform recommendations for improving CN 3.0.

Additional Considerations for Constituent Voice: The concept of constituent voice is a fairly unique metric in the nonprofit ratings and standards field. Therefore, the SPEA Capstone Group had to vary its approach to learn about this element. The main source of background information came from a report entitled *21st Century Potential of Constituency Voice*:

Opportunities for Reform in the United States Human Services Sector, which is co-authored by Keystone Accountability, a partner to Charity Navigator. Though the Capstone Group conducted academic research about similar topics like “feedback” and “client satisfaction,” most of the existing literature does not apply to the comprehensive concept of constituent voice, as understood by CN.

Charity Data Collection

Charity Navigator 3.0 Research Tool. Charity Navigator identified a sample of 90 organizations in its data base in three cause areas: Homeless Services, Children’s and Family Services, and Social Services. Each student was assigned six to eight charities to rate, using the CN 3.0 Research Tool (CNRT); each charity was rated by two students to create a basis for inter-rater comparisons. One hundred and seventy-four responses were collected using the tool. (One organization’s website did not work. Four organizations were rated by just one student.) Students were instructed to read the Charity Navigator instructions before beginning rating charities, and several discussions occurred during class meetings about the importance of following the instructions. Though the Capstone Group has no quantitative data on the time students spent reading the instructions, many journals referenced referring back to the instructions during the rating process.

The CN 3.0 Research Tool was designed to capture online a beginning and ending time for rating each charity. However, there was a flaw in the tool design that resulted in no ending timestamp if the student answered “No” to the first test of Element Three. Additionally, upon analyzing data generated by the tool, it appears the ending timestamp did not take into account the researcher’s time zone (which could vary depending on when and where the student used

the tool). Finally, there were inconsistencies in reporting as the timestamp was based on a 24 hour clock, which some students misinterpreted. Since time was recorded within CNRT itself, the SPEA Capstone Group could not correct mistakes. As a result, the Capstone Group concluded that much of the data on the time needed to apply the tool was unreliable and did not analyze it.

Capstone Background Research Tool: To gain additional understanding of the amount and quality of information related to results reporting that is currently published on charities' websites, the Capstone Group initially developed a Background Research Tool (BRT) to guide data collection by students based on the first concept note provided by Charity Navigator. Students identified types of results-reporting information that might be related to each of the original three elements and cross-referenced their list with best practice research and the CN 3.0 prototype rating system documents to produce a comprehensive list of potential measures. In an effort to ensure that researchers would collect the same information as a group, clarifying questions for each item were included to guide the student researchers as they reviewed each website. Students chose three to four charities from the original six to eight assigned them and performed an assessment of the charities' website using the BRT. They collected 89 responses in total.

Students spent an average of 43 minutes using the BRT per charity. The most frequently reported time (mode) was 30 minutes. The range was from 20 to 75 minutes per charity. Since the BRT was both more comprehensive than and different from the CNRT, these figures are not necessarily a guide to the time-to-completion for the CNRT.

Student Journals: Students recorded qualitative information in journals about their experiences with the Tool and the Background Research Guide. Every student wrote one journal entry for each charity rating with the CNRT and one entry for each BRT collection. A total of 174 CNRT journals and 89 BRT journals were collected.

Charity Data Analysis

Frequency Tables: The data from the CN 3.0 Research Tool was synthesized into frequency tables, organized in each cause area, and subsequently reviewed as a group. In addition, students calculated the percentage of “Yes” or “No” responses for each cause area based upon the total number of observations in each cause area (60 each for Social Services and Children’s and Family Services, and 54 for Homeless Services), and then calculated total percentages out of the 174 responses collected. Finally, students rank-ordered the data by the highest percentage of “Yes” responses for each category and also for the aggregate, which was comprised of all three cause areas.

Inter-Rater Reliability: To assess the consistency between researchers, or inter-rater reliability, the Capstone Group separated the CNRT data into pairs of charities and identified all concordant and discordant pairs (i.e., where both raters agreed or disagreed). These pairs were counted and percentages were presented by cause area and in the aggregate.

For discordant pairs on the compliance footprint question, an average spread was also generated by looking at the average disparity between the estimates of the alignment of mission and resources.

Discordant pairs that had researchers listing either one numerical and one non-numerical rating or two non-numerical ratings were also tabulated. Percentages are presented by cause area and in the aggregate.

For researcher confidence level, discordant pairs were identified and counted. Percentages were presented by average spread and cause area, and an average of the differences between researchers was generated.

The number of concordant and discordant pairs is important to note because it is one way of assessing the reliability of the CNRT. If the CNRT presents different results based on who is rating a charity, the assessment of a charity is not as objective as is desired. Discordant pairs on the compliance footprint reveal differences in rater cognition of a charity's alignment of resources and mission. High variability between researchers, evidenced through the magnitude of the average spread, potentially indicates an insufficient pre-rating briefing by the rater and can be addressed through more detailed rating guidelines and directions. Discordant pairs presenting a numerical and non-numerical rating, or two differing non-numerical ratings, are particularly problematic because researchers working with the same information are producing significantly different ratings. In terms of researcher confidence level, a large number of discordant pairs show that most researchers disagreed in making their decisions.

Student Journals: The Capstone Group reviewed all 176 journals, gleaning key concerns and major themes commonly mentioned about using the CNRT. Discussion groups of four to five students, one from each team if possible, further analyzed key themes, including CNRT instructions and question structure, variance among charities, compliance footprint, standards

and accreditation organizations, and evaluation and assessment. The information gleaned through this process was used in formulating recommendations for Charity Navigator.

Additional Considerations for Constituent Voice: Data from the CNRT indicates most nonprofits in the sample were either unaware of the constituent voice concept or did not emphasize it on their websites. Only a small amount of data from the CNRT (just 31 out of the original 174 responses) met the threshold question of the first constituent voice test. Analysis of this element, therefore, is more limited because of sample size than Elements One, Two, and Four, making it more difficult to extrapolate conclusions about the Tool. Despite this limitation, Capstone Group feels that its conclusions and recommendations for the CNRT are valid for the 31 responses provided. However, assessing inter-rater reliability has little utility, considering that only four pairs of raters provided a concordant “Yes” response to the first test, the threshold question.

Charity Navigator Constituent Feedback Collection

Charity Phone Survey: Supplemental phone surveys were conducted with sample of charities’ leaders to allow students to gain qualitative feedback about the utility of the CN 3.0 framework from the charities’ perspective. Responses to the CN telephone survey and supplemental student-created questions were recorded in telephone journals by each student. Phone surveys were successful in producing responses for 35 of the 90 charities. This reflects a smaller sample size than desired, but also the reality of working with busy people in the nonprofit sector.

Online Member Survey: Students created and implemented an online survey for CN members to gather feedback about the utility of the CN 3.0 framework from the members’

perspective. SurveyMonkey was used as the platform for design and launch of the survey due to cost and technological considerations. The survey contained five questions, each with a section for optional free response comments. A pre-test of 20 non-CN member respondents was conducted to test the survey for any bias or confusion; edits were made according to the resulting suggestions. The SurveyMonkey link was sent via email to a listing of CN donors who gave \$25 or less, 3,731 in total, and was active for one week (March 12 – 19, 2012). A total of 536 respondents started the survey, with 489 completing it by the expiration date. The total number of respondents was 536, which indicates a 15 percent response rate (Appendix F presents survey questions).

Charity Navigator Constituent Feedback Analysis

Charity Phone Survey: Student raters followed a script to approach and interview charities. The phone surveys contained four questions written by CN and four supplementary questions written by Capstone Group Team Five. Student raters were encouraged to ask as many questions as possible in the time allotted by the charity and to record their responses in journals. The 54 charities not represented in the phone survey journals either declined to answer survey questions, could not be contacted, were contacted but did not follow up on messages left by student raters, or otherwise did not follow through on student raters' contacts. The resulting journals of the 36 charities providing answers to phone survey questions were coded for relevant themes.

Online Member Survey. Though the Capstone Group survey tried to be as methodologically sound as possible, two critical issues may affect the results. First, the results can only be generalized to the broader donor population with caution since the sample was

drawn from CN members, who may or may not be typical of the larger group (or even of CN website visitors) in knowledge of or desire to see more results reporting by charities. A series of questions to determine demographics would have eased this potential problem and should be included in any future surveys.

Second, the survey assumed that CN members would have prior knowledge of the concepts included in the CNRT. In fact, the survey should have been designed so that respondents without an appropriate knowledge of results measures, evaluations, and constituency voice would have been screened out and prevented from answering attitude questions. A passage from *Asking Questions: the Definitive Guide to Questionnaire Design* explains this:

Knowledge questions help qualify respondent opinions and should be asked before attitude questions are asked. This order is essential if the knowledge questions are to screen out respondents who do not have sufficient information to answer detailed attitude questions. If the attitude questions are asked first, respondents may feel they are expected to know about the issue and have an opinion (Bradburn, Sudman, and Wansink, 2004).

There is evidence that respondents were prompted by the concepts presented in the survey to give an opinion though they lacked sufficient knowledge. Language conveying lack of concept knowledge, evidenced by words such as “should” or “future”, was used throughout the free response comment sections. (“Should” and “future” were used in 20 of 488 free response comments, or 4%). Since free response comments were optional, the significance of this kind of response is not clear.

AmeriCorps Case Study

Later in the report, the Capstone Group Team Six will discuss the application of the CN 3.0 Elements for improving the AmeriCorps Improving Health Throughout Indiana's (AIHTI) evaluation process. This was also a test of how the CN concepts could be applied to a government-supported program. The case study will discuss how Capstone Group Team Six applied the Charity Navigator methodology and framework to a set of programs outside the scope of Charity Navigator's ranking system. It will also explore ways that AIHTI can implement the CN 3.0 standards in order to increase its effectiveness.

Chapter Three

Charity Information Intermediaries Review

What is a Charity Information Intermediary?

Information intermediary organizations exist to monitor the nonprofit sector beyond the legal reporting requirements to federal and state governments. In a 2005 study, “Rating the Raters,” intermediaries are defined as organizations that are “national in scope or being replicated across the country” that “rated, ranked, or set standards for diverse CNOs [charitable nonprofit organizations] rather than a specialized niche groups of nonprofits” (National Council of Nonprofit Associations and National Human Services Assembly, 2005, p. 5).

Two main types of organizations exist in the charity accountability field: raters with point- or star-systems and non-raters with standards and guidelines. Charity Navigator (2012) describes itself as a ratings organization designed to “help charitable givers make intelligent giving decisions by providing information on over five thousand charities and by evaluating the financial health of each of these charities,” with the goal “to advance a more efficient and responsive philanthropic marketplace.” Another intermediary agency describes itself as a “national initiative to promote the highest standards of ethics and accountability in nonprofit governance, management and operations, to facilitate adherence to those standards by all nonprofit organizations,” with no mention of ratings (Standards for Excellence Institute, 2004, p. 32). Both types of intermediaries have charity effectiveness and accountability as the end goal, but each promulgates its standards differently.

Locally-based intermediaries also exist and have the same split methodology.

Minnesota's Charities Review Council is an independent 501(c)(3) organization that rates and grants seals of approval to public charities. Many states, including Maryland and Minnesota, also have statewide associations of nonprofits that create guidelines for implementing best practices in effectiveness; however, these organizations do not certify or rate charities.

Nonprofit Accountability and Effectiveness

As accountability literature from the 1990s and 2000s indicates, concern over nonprofit performance, results, governance, and fiduciary practices is not a new phenomenon (Ebrahim, 2010). While Kearns (2001) points out common law regulation of charitable trusts in England going back as far as the 16th century, Light (2002) highlights the impact of nonprofits' behavior about donations in the wake of the September 11th attacks as a modern impetus for efforts at self-regulation. Independent Sector's Panel on the Nonprofit Sector was convened in 2004 in response to calls from the U.S. Senate Finance Committee (Panel on the Nonprofit Sector, 2007). Murray (2010) attributes the government and public interest in nonprofit accountability to the U.S. government's efforts to make its agencies more efficient and responsive. Together, these pushes toward nonprofit organizational effectiveness and efficiency measures are part of what Murray (2010) calls the "accountability movement" (p. 434).

Alnoor Ebrahim's (2010) analysis of nonprofit accountability offers four functional focus areas: Finances, Governance, Performance, and Mission. Nonprofits undergoing an accountability assessment should expect to address each of these categories in some fashion, regardless of the type of intermediary (rater or standards). Charity Navigator and other intermediaries currently assess finances and governance to a greater extent than mission and

performance. Data on finances and governance are more easily accessible through Form 990, accessed electronically through the Internal Revenue Service (IRS). CN 3.0 is Charity Navigator's effort to move charities toward reporting more performance data. In later sections, the Capstone Group outlines other intermediaries' attempts to encourage results reporting.

Criticisms of Information Intermediaries

Some nonprofit scholars and practitioners believe intermediaries present problems for the charitable sector.

Mission Displacement: Many practitioners and opinion leaders in the sector disagree with the goals of intermediaries—especially ratings organizations. Nonprofit scholar Bruce Hopkins (2009) defines intermediary agencies as “nonprofit, tax-exempt, charitable organizations, established to tell other nonprofit, tax-exempt (usually), charitable entities how to operate and punish them if they do not adhere to the intermediaries' dictates (which are often inconsistent)” (p. 257). Some believe the risks to the sector presented by placing emphasis on performance and results are large, including a potential for goal displacement, bias in evaluation methods, and encouraging standardization, rather than innovation and adaptation (Gronbjerg, 2012).

Multiple Constituents: J. Stephen Ott (2001) highlights another problem with nonprofit accountability: “*Accountability* means *answerability* for one's actions or behavior. The more difficult questions for nonprofit organizations are: accountable to whom and accountable for what?” (p. 345). Because the staff, board, clients, donors, government partners, and other constituents of each nonprofit have a variety of perceptions and needs, each nonprofit may face conflicting standards of efficacy and efficiency. Adding outside information intermediary

standards and ratings mechanisms to the list further complicates the reporting demands placed on the charity.

Comparison Information Intermediaries

To understand the landscape of the charity information intermediary field, the Capstone Group drew comparisons between Charity Navigator's standards and ratings system and that of ten other comparable intermediary organizations. (Appendix B gives a detailed description of each of them.) It began research with an environmental scan and selection of 19 charity information intermediaries, judged the most significant:

- The Better Business Bureau's Wise Giving Alliance
- Charities Review Council
- Charity Watch (American Institute of Philanthropy)
- Charting Impact (partnership of Independent Sector, The Better Business Bureau's Wise Giving Alliance, and GuideStar)
- ECFA
- GiveWell
- GreatNonprofits
- GuideStar
- Independent Charities of America
- InterAction
- Intelligent Philanthropy
- MinistryWatch
- Panel on the Nonprofit Sector (convened by Independent Sector)
- Philanthropedia (recently acquired by GuideStar)
- Standards for Excellence Institute
- The Charity Rater
- Partners for Change

- Root Cause

From these 19, ten were identified as comparable to Charity Navigator, based on similarities in their mission, methodology, assessment mechanism, and summary strategy.

The sample of ten includes the following:

- The Better Business Bureau's Wise Giving Alliance
- Charities Review Council
- CharityWatch
- ECFA
- GiveWell
- GuideStar
- Independent Charities of America
- Intelligent Philanthropy
- InterAction
- Standards for Excellence Institute

Though the Capstone Group did not include it in the sample because its activity is relatively recent, the Urban Institute is included in Appendix B because of its work in the results reporting field.

Though each intermediary takes its own approach to assessing efficiency and effectiveness in charities, the benefits of looking at them as a group outweigh the challenges. First, Charity Navigator, while perhaps the best-known intermediary in its field, does not operate in a vacuum. Charities may experience pressure to get the "seal of approval" from more than one intermediary; for example, a common observation during the research phase of the Capstone Group project was a Better Business Bureau and Charity Navigator seal side-by-side on a charity's homepage. The degree to which Charity Navigator's rating system differs

from other intermediaries will directly correlate to the amount of effort charities must exert to meet more than one intermediary's requirements for approval.

In addition, the Capstone Group compared CN 3.0 Elements to other intermediaries to test how familiar the concepts are likely to be. If similar components are not being used by other intermediaries, Charity Navigator will need to take steps to ensure that both the charities it is rating and users of its site understand the kinds of results reporting practices it is seeking to encourage.

Points of Comparison

Mission: The intermediaries analyzed have similar purposes for their indicators. Most intermediaries aim to increase public trust and nonprofit funding. Three of the intermediaries, Intelligent Philanthropy, Better Business Bureau, and Charities Review Council, aim to increase donor knowledge and awareness, and three explicitly state that they aim to increase nonprofit effectiveness (Standards for Excellence Institute, Better Business Bureau, GuideStar and its affiliates Charting Impact and Philanthropedia). Interaction, Better Business Bureau, and Intelligent Philanthropy also aim to provide recognition to successful organizations. Additional intermediary purposes include the following: increasing policy awareness and addressing potential policy changes, empowering donors to discover answers on their own, and encouraging fair, honest, and ethical solicitation practices.

Methodology: Intermediaries use various resources to gauge whether charities are meeting the requirements of their ratings, including the following: bylaws, articles of incorporation, letters of recommendation, written reports, analytical overviews (Intelligent Philanthropy), site visits by experienced professionals, field and/or compliance reviews, phone

calls, internal charity documents, analysis of programs, mission statements, annual Form 990, and strategic plans. Those intermediaries that do not rate charities provide online sources and other documents for charities and donors to use to rate charity effectiveness or commitment to determining effectiveness, such as GuideStar's "Take Action" page (GuideStar, 2012).

Assessment Mechanism: The types of raters and assessors of charities vary among these intermediaries, as does the frequency with which reassessment is performed. Most assess charities against standards using the intermediary's staff. This mechanism constrains the number of charities these intermediaries can rate, a problem Charity Navigator is working to overcome in CN 3.0 by using non-staff raters. InterAction uses a committee made up of members from its Board of Directors. GuideStar allows charities to submit information until they have met its requirements. Intelligent Philanthropy and The Better Business Bureau also allow rated charities to resubmit information annually.

Summary Strategy: Two of the ten (GiveWell and InterAction) do not offer seals or recognition to charities. These intermediaries list endorsed charities on their website. Eight of the ten intermediaries analyzed offer some type of seal or recognition badge for meeting its requirements. Table 3.1 shows seals or certifications offered to charities that are in compliance with intermediaries' indicators.

Table 3.1: Seals or Recognition Summary Measures

Intermediary	Seal or Recognition Badge	Levels of Recognition	Additional Information
Better Business Bureau Wise Giving Alliance	National Charity Seal	"Accredited Charity" - Meets 20 Standards "Does Not Meet One or More Standards" "Nondisclosure" - Did not provide requested information	Accredited Charities must pay a fee (\$1,000 to \$15,000 depending on size) to display the Seal
Charities Review Council	Meet Standards Seal	Pass or Fail	
Charity Navigator	4-Star Charity Navigator Logo	0 - 4 Stars indicating: 0 - Exceptionally Poor 1 - Poor 2 - Needs Improvement 3 - Good 4 - Exceptional No rating indicates "Donor Advisory"	Only 4-Star Charities given option to display the Logo Charities with a pattern of excellence in ratings periodically publicized on "Top Ten" lists
CharityWatch	Letter Grade Seal Top-Rated Seal	Letter grade indicating: A - Excellent B - Good C - Satisfactory D - Unsatisfactory F - Poor ? Indicates insufficient information	A+ to B+ organizations publicized as "Top-Rated"
ECFA	ECFA Seal	Pass or Fail	Charities that meet the Standards must pay a membership fee (\$525 to \$10,500 depending on size) to display the Seal
GuideStar	Guidestar Exchange Seal	Pass or Fail	
Independent Charities of America	Seal of Excellence	Pass or Fail	
Intelligent Philanthropy	Commitment to Transparency Seal	Pass or Fail	
Standards for Excellence Institute	Seal of Excellence	Three Tier System indicating: "Essential" - Adherence to Basic Legal, Regulatory, and Governance Practices "Enhanced" - Enhancing the Foundations of Nonprofit Management and Governance "Certified" - Meets all 55 standards	Only Certified Charities may display the Seal

Source: Intermediaries' Websites, 2012.

Elements Used by Information Intermediaries

Logic, Results, and Measures:

CN 3.0's results logic and measures element has two key defining components:

1. How the charity's activities are intended to lead to pre-defined outputs and outcomes
2. How results-related data collected by charities are meant to mark progress against those results (Charity Navigator 3.0 Research Guidelines, 2012).

Element One shares some similarities with the components of results measurement that are included in seven of the ten intermediaries' indicators. (See Table 1.3.) In relation to CN 3.0's overall definition, both CN and other intermediaries assess whether charities define how outputs may lead to defined outcomes and whether charities track progress towards meeting outcomes. CN 3.0, however, breaks down these defining components into specific statements that are to be observed on a charity's website. Users of the Charity Navigator 3.0 Research Tool are expected to look for a clear statement or set of statements that indicate a charity's commitment to these initiatives.

The Logic, Results, and Measures element, as described in CN 3.0, is not a prominent indicator utilized by any of the ten Charity Navigator comparison intermediaries. Components of results measurement are incorporated sporadically within or as supplements to other indicators. However, no intermediary utilizes the element in its entirety.

Generally, intermediaries that include a type of results measurement emphasize "demonstrated impact" as an essential component of this element. However, the concept of demonstrating impact varies across intermediaries.

“Demonstrated impact” is usually defined as incorporating a policy or procedure to determine results. Some intermediaries define demonstrating impact as charities having procedures for collecting results-related data or descriptions of accomplished goals that are aligned with a charity’s mission. Only three of these intermediaries specifically state the need for charities to have procedures for collecting quantitative and qualitative data. For example, Standards for Excellence Institute rates charities on having procedures for assessing program effectiveness. InterAction requires that charities have both procedures for assessing mission and operation effectiveness through ongoing strategic planning and defined procedures for evaluating programs based on both qualitative and quantitative measures. Charities Review Council rates whether charities describe accomplishments from the past year in relation to their missions and future goals for the next year.

Overall, four of the seven intermediaries that include results measurement (InterAction, Better Business Bureau’s Wise Giving Alliance, Standards for Excellence Institute, and Charities Review Council) require charities to have a policy or procedure in place to assess effectiveness or their program impact in fulfilling goals, whereas other intermediaries provide frameworks or forums to assist charities in assessing and communicating impact. For example, three of the seven organizations, including GuideStar and Intelligent Philanthropy, provide support or forums to aid charities in tracking progress towards outcomes. GiveWell is the only intermediary that focuses on how successful programs actually are (GiveWell, 2012).

Several intermediaries consider communication of program goals and mission alignment with those goals among results logic and measures indicators. For example, The Better Business Bureau asks that charities define measurable goals and objectives, as well as assess

effectiveness in achieving them in relation to the charity’s overall mission. Pre-defined goal and tracking-strategy requirements are similar to CN 3.0’s idea of “pre-defined outputs and outcomes” (Charity Navigator 3.0 Research Guidelines, 2012).

Additionally, a common defining component of this indicator is whether or not a charity indicates it tracks progress towards goals to assess short-term impact. Six intermediaries refer to tracking progress as a key component of results measurement. Of the seven intermediaries that include some form of results measurement, five include a statement regarding the need for defined program goals, as well as strategies for reaching those goals, within their results measurement indicator. Tracking progress is similar to CN 3.0’s statement that charities should “mark progress against results.” The Charity Navigator 3.0 Research Tool, however, requires charities to state how outcome-related data collection is meant to mark that progress. (See Table 3.2 for an illustration of how other information intermediaries use the results logic and measures element.)

Table 3.2: Illustration of How Comparison Organizations Use Logic, Results, and Measures in Charity Assessment

CATEGORY	INTERMEDIARY	
Results as Core Element	GiveWell (rating actual impact) BBB (rating whether or not policies exist to assess impact) SEC (Standard 1: Mission and Programs)	50%
Results Question one of a few indicators	GuideStar	17%
Results Question or Standard included within other standards (not separate as an indicator)	InterAction (within standard 6) CRC (in public disclosure strand)	33%

Independent Evaluations:

CN 3.0's independent evaluation and standards and certification mechanisms element has two key defining components:

1. Whether the charity publishes rigorous independent evaluations of its programs
2. Whether the charity subscribes to relevant standards/certification bodies that take into account outcome measurement (Charity Navigator 3.0 Research Guidelines, 2012).

Other intermediaries address components of this element to various degrees. No other intermediary specifically takes into account *independent* evaluations as Charity Navigator does; however, six other intermediaries utilize an evaluation indicator of some sort. (See Table 1.3.) Furthermore, no other intermediary takes into account additional standards/certification mechanisms as CN 3.0 does.

Although many other intermediaries take internal program evaluation and benchmarking into account, there is currently very little focus on solely independent evaluations among these intermediaries. Among intermediaries that take into account evaluations in any form, there exists little cross-organization agreement within the field regarding the best definition or terminology for this indicator. Examples of phrases and concepts other intermediaries use include the following: evaluating “performance and effectiveness” (Better Business Bureau, 2003), reviewing the charity’s mission annually in regards to community impact (Charities Review Council, 2010), providing self-evaluation questions for leadership to ask (Charting Impact, n.d.), and inclusion of both quantitative and qualitative data in “candid” evaluations (Standards for Excellence Institute, 2009).

Intermediaries such as the Standards for Excellence Institute recommend that a charity should have a specific policy regarding evaluation, but do not specify parameters for the evaluation process (Standards for Excellence Institute, 2009). The intermediary rationale behind inclusion of an evaluation indicator centers predominantly on the importance of an organization's being able to determine effectiveness. Closely tied to this rationale is the importance of donors' being able to determine where their donation could make the greatest impact.

CN 3.0's definition of independent evaluations is very thorough compared to other intermediaries. It states in its CN 3.0 research instructions that charities "get credit for having rigorous independent evaluations of their programs" and that charities themselves must report an "evaluation quality rating" for published evaluations (Charity Navigator, 2012). No other intermediaries attempt to examine the proportion of charity programming covered by evaluations. Furthermore, no other intermediaries require a charity to report a rating of the quality of the evaluations. The clear conclusion from comparing CN 3.0's definition of this indicator to that of other intermediaries is that CN 3.0 is concerned with solely *independent* evaluations in this element, while other intermediaries have a broader definition of evaluation that does not take into account the independence of an evaluation, but rather recommends or requires an evaluation of any sort to take place periodically (including completely internal evaluations).

Although CN 3.0 is leading the way in the inclusion of independent evaluations as an indicator of effectiveness, other intermediaries do recognize the relevance of evaluation in gauging charity effectiveness. It is important to emphasize again that the indicators used by other intermediaries, however, *do not* differentiate internal and external evaluations, or

require that an evaluation be independent. Many of the intermediaries that fail to include an evaluation indicator often include an indicator for results reporting, but for this examination of CN 3.0, the Capstone Group did not consider a results reporting requirement comparable to a rigorous independent evaluation requirement.

Similarly, CN 3.0's inclusion of additional reporting standards and certification mechanisms is also highly innovative. No other intermediary researched by the SPEA Capstone Group takes into account additional standards and certification mechanisms; furthermore, these other intermediaries do not include additional standards and certification mechanisms that emphasize outcome measurement as CN 3.0 does. Although Charity Navigator's inclusion of charity participation in these reporting standards and certification regimes is unique and valuable, it is complicated by the volume and variety of standards and certifications in the non-profit sector. Such standards and certifications provide charities with public benefits, including branding opportunities (such as displaying a seal or certification on a website), as well as private benefits, including increased organizational learning and opportunities for improved effectiveness (Gugerty and Prakash, 2010). As implementing CN 3.0 is a multi-year effort, creating a comprehensive database of relevant standards and certification mechanisms should be done and will certainly help the rater conduct the rating process within a specified amount of time.

Independent Sector has compiled a list of nearly 100 principles, standards, and codes for nonprofit and philanthropic organizations in its "Compendium of Standards, Codes, and Principles of Nonprofit and Philanthropic Organizations" (Independent Sector, 2012). This list includes industry-specific organizations, such as the American Hospital Association and the

National Health Council, but many of the organizations do not include outcome-based criteria, as CN 3.0 does. Standards and certification mechanisms in the charitable sector are highly complex with variations in geographic scope, industry focus, and the degree to which outcomes are included.

Constituent Voice:

CN 3.0's constituent voice element has two key defining components:

1. Whether the charity collects and publishes feedback from its primary constituents
2. How well the charity collects and publishes feedback from its primary constituents (Charity Navigator 3.0 Research Guidelines, 2012).

No other intermediary considers this element in assessing charities' effectiveness, and therefore, no agreed-upon definition of "constituent voice" exists outside of Charity Navigator's conception. Charity Navigator, in cooperation with Keystone Accountability, has created the constituent voice tool that provides detailed guidelines to assess charities' efforts to collect and publish feedback.

A charity has many constituencies including clients, donors, staff, volunteers, neighbors, and all "who have a significant influence on a set of social issues, both as they are now and as they are envisioned in the future" (Keystone Accountability, 2012). "Constituent voice" is defined as "the practice of ensuring that the views of all relevant constituents are seriously taken into account in planning, monitoring, assessing, reporting and learning from an

organization’s work” (Keystone Accountability, 2012). CN 3.0’s inclusion of the constituent voice element is highly innovative in comparison with other intermediaries.

Furthermore, the published feedback must come from primary constituents, namely “people who are meant to be the direct recipients of benefits created by the charity’s actions” (Charity Navigator 3.0 Research Guidelines, 2012). Many people, such as the donors and partners, can be considered non-primary constituents since they benefit indirectly.

Alignment of Mission and Resources:

CN 3.0’s alignment of mission and resources element has one key defining component: Whether the charity uses its financial resources to support mission-related activities (CN Element Four Definition, 2012).

Charities that pass Element Four’s criterion have shown that its mission and program work are in line with their finances and staff time (CN Element Four Definition, 2012). Public claims regarding a charity’s work are compared to a charity’s financial reports to determine whether the charity’s claims about resource allocation match its Form 990’s reporting of program expenses. Element Four is based upon how a “reasonable person” would assess the alignment of charity claims and use of its resources (CN Element Four Definition, 2012).

Four Elements in Practice:

Other information Intermediaries utilize components of the CN 3.0 elements as parts of their indicators. However, the degree to which they use components that are similar to Charity Navigator’s four elements varies. Table 3.3 compares the sample intermediaries’ use of concepts related to CN 3.0’s Four Elements.

Table 3.3: Comparison of Ten Intermediaries' Use of CN 3.0 Four Elements Concepts in Ratings and Standards Systems

Intermediary	CRITERIA FOR ELEMENT 1	CRITERIA FOR ELEMENT 2			CRITERIA FOR ELEMENT 3	CRITERIA FOR ELEMENT 4
	Results Logic & Measure Indicator	Evaluation Indicator	Independent Evaluation Indicator	Standards/Certification Indicator	Constituency Voice Indicator	Compliance Footprint Indicator
Better Business Bureau's Wise Giving Alliance	Yes	Yes	No	No	No	No
Charities Review Council	Yes	Yes	No	No	No	No
CharityWatch	No	No	No	No	No	Yes
Evangelical Council for Financial Accountability	No	No	No	No	No	No
GiveWell	Yes	Yes	Yes (as selection mechanism for charities)	No	No	No
GuideStar	Yes	Yes (as part of Charting Impact partnership)	No	No	No	No
Independent Charities of America	No	No	No	No	No	No
Intelligent Philanthropy	Yes	No	No	No	No	No
InterAction	Yes	Yes	No	No	No	No
Standards for Excellence Institute	Yes	Yes	No	No	No	No

Logic, Results, and Measures: Of the ten intermediaries analyzed, seven of them refer to results measurement: The Better Business Bureau Wise Giving Alliance, Charities Review Council, GiveWell, GuideStar, Intelligent Philanthropy, InterAction, and Standards for Excellence Institute (See Table 3.3). However, two of the seven do not actually rate charities and the degree to which results measurement is important to intermediary approval ranges from being

a core element to being a sub-question among several other indicators. Four intermediaries—Give Well, Better Business Bureau, Standards for Excellence Institute, and Intelligent Philanthropy—include results measurement as a core element of their rating system. Intermediaries like GuideStar include results logic and measures more tangentially as an optional component and InterAction and Charities Review Council include it as sub-questions within other major indicators.

Independent Evaluations: Of the ten intermediaries analyzed, only one refers specifically to independent evaluations, and does so as a *selection* mechanism for charities to be rated, not as a *rating* mechanism itself. However, six intermediaries refer to evaluations without specifying the required degree of independence. Among these six intermediaries, there is considerable variation regarding the degree to which evaluations are required or simply recommended or encouraged. Two intermediaries require evaluation in some capacity (the Better Business Bureau and Charities Review Council), while three intermediaries recommend or encourage evaluation (GuideStar [through its Charting Impact self-evaluation partnership], InterAction, and Standards for Excellence Institute). The remaining intermediary (GiveWell) utilizes independent evaluations as a selection mechanism as opposed to a rating mechanism. Finally, no intermediaries take into account additional standards and certifications.

Constituent Voice: Of the ten intermediaries analyzed, no intermediary addresses the constituent voice element.

Alignment of Mission and Resources: Intermediaries are currently not assessing charities in this way. CharityWatch includes an indicator known as “Percent Spent on Charitable Purpose” which shares some similarities with Element Four. CharityWatch utilizes

Form 990, in addition to audited financial statements and annual reports, to determine total expenses spent on charitable programs. CharityWatch claims to provide financial adjustments to better inform donors seeking efficient use of their funds (CharityWatch, 2012). However, CharityWatch rates the charity on this indicator using its own standards, unlike CN, which is focused on how “any reasonable person” would judge charity claims regarding the alignment of their work and their finances (CN Element Four Definition, 2012). Additionally, CharityWatch states that charities should spend at least 60 percent of total expenses on programming, with the most efficient closer to 75 percent or more (CharityWatch, 2012). CN 3.0’s Element Four does not state a particular percentage of funds that should be used for programming, but focuses comparing what the financial reports say and what a charity claims to its constituents. Overall, the information intermediaries analyzed in this report are not using Element Four, nor does other research describe this indicator as CN does.

Conclusion

These comparisons show key similarities and differences Charity Navigator’s Four Elements and the rest of the information intermediary field. Overall, Charity Navigator’s 3.0 concepts are different from other organizations’ ratings systems. The results logic and measures element is fairly prevalent already. Many intermediaries encourage evaluation and program assessment, but not as CN 3.0 seeks (independent versus internal or collaborative evaluation). Virtually no other intermediaries are assessing charities on the Constituent Voice or Alignment of Mission and Resources elements.

If Charity Navigator develops a solid understanding of how its new rating system fits into the larger field of information intermediaries, it can use this information to differentiate itself

and promote its ratings philosophy. Rated charities will also benefit from Charity Navigator's knowledge of the field. These nonprofit organizations often feel pressure to get more than one intermediary seal or recognition, so Charity Navigator may want to be sensitive to the burdens that different standards put on rated charities.

Chapter Four

Element One – Logic, Results, and Measures

The goal of Element One is to assess the causal link between a charity's activities and how they contribute to pre-defined outputs and outcomes, and to ascertain if a charity is using results-reporting data to track its progress and overall impact. While it is clear from the Capstone Group's research that some charities have implemented sophisticated performance measurement systems and practices, a great number of them have almost no data-driven basis for knowing that the programs or services they provide are making any impact. Charity Navigator has developed Element One in the hopes of incentivizing charities to publish any efforts regarding this important work. Overall, the Capstone Group agrees that Element One's tests provide a good initial framework. Nevertheless, it believes that expanding each test's scope in order to gather data for each concept more comprehensively can enhance the CNRT.

The following section will discuss the findings from conducting research using the Charity Navigator 3.0 Research Tool. Those findings will then be used as a point of comparison for the research that the Capstone Group completed using the Background Research Tool. Subsequent chapters will include recommendations and a discussion about the limitations of this research.

The Four Tests for Element One

Element One is comprised of four tests. Each test includes the following:

- One main binary question

- A question related to compliance footprint which asks researchers to choose from a list of percentage scores
- A question related to the researcher's confidence level which is rated on a three-point scale
- An open-ended response question that prompts researchers to list the main challenges in feeling confident about their answers

The CNRT questions regarding Element One intend to evaluate the quality and rigor of a charity's published results-reporting information.

Test One asks, "Is the causal logic plausible?" The purpose of this test is to assess if the programs and services the charity provides are likely to lead to pre-defined outputs and outcomes. Test Two asks, "Is there some indication of how much of the action is required to produce what effects (i.e., dosage)?" This test is two-pronged in that it is looking for both a clear statement of how much of the programs and services are needed to produce desired outcomes, and the outcomes themselves. Test Three asks, "Is there some indication that the logical inference (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)?" This test asks if the charity links its actions to proven methods through reference to published evidence. Test Four asks, "Are there specified measures (indicators) to be collected and a plan to do so?" This test assesses whether or not the charity collects data systematically on indicators of results for its specific programs and services.

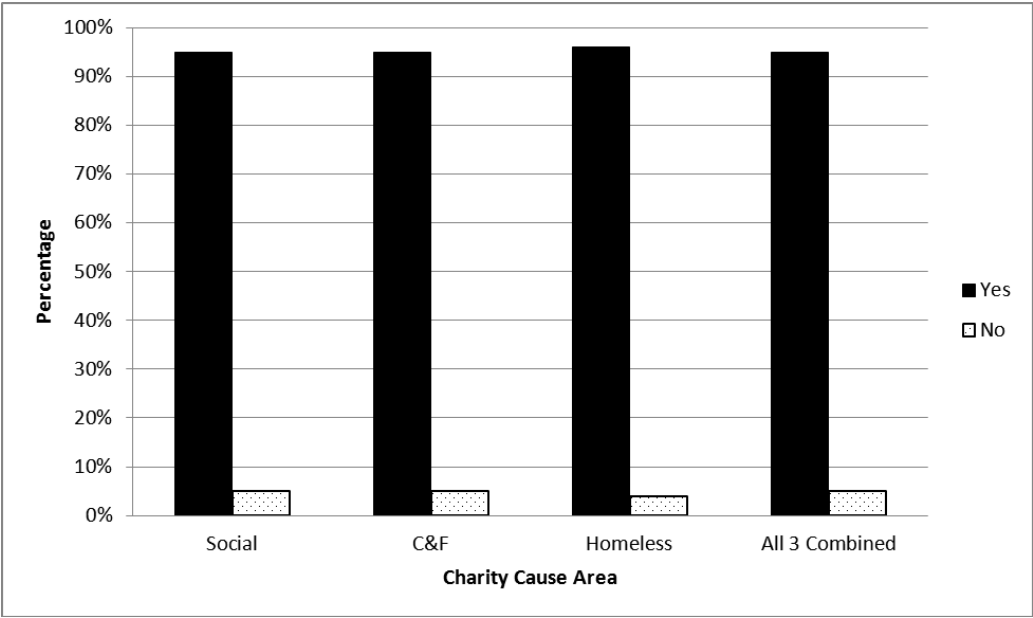
Students filled out 174 surveys for the CNRT. These surveys covered charities within three cause areas, so the percentages provided below for all three cause areas combined (Social Services, Children's and Family Services, and Homeless Services) are out of a total of 174.

The percentages provided for Social Services and Children’s and Family Services organizations are both out of a total of 60 student-completed online surveys, and the percentages provided for Homeless Services organizations are out of a total of 54 student-completed online surveys. The binary question and researcher confidence level results for the four tests of the Results, Logic, and Measures Element are summarized in the Frequency Tables. (See Appendix C). Student responses to open-ended questions in the CNRT were analyzed to understand the challenges that students confronted while completing the Tool. Because two students reviewed each charity to determine if there would be differences in quality of research and agreement, analyses on inter-rater reliability for each Test are also provided.

Test One: Is the Causal Logic Plausible?

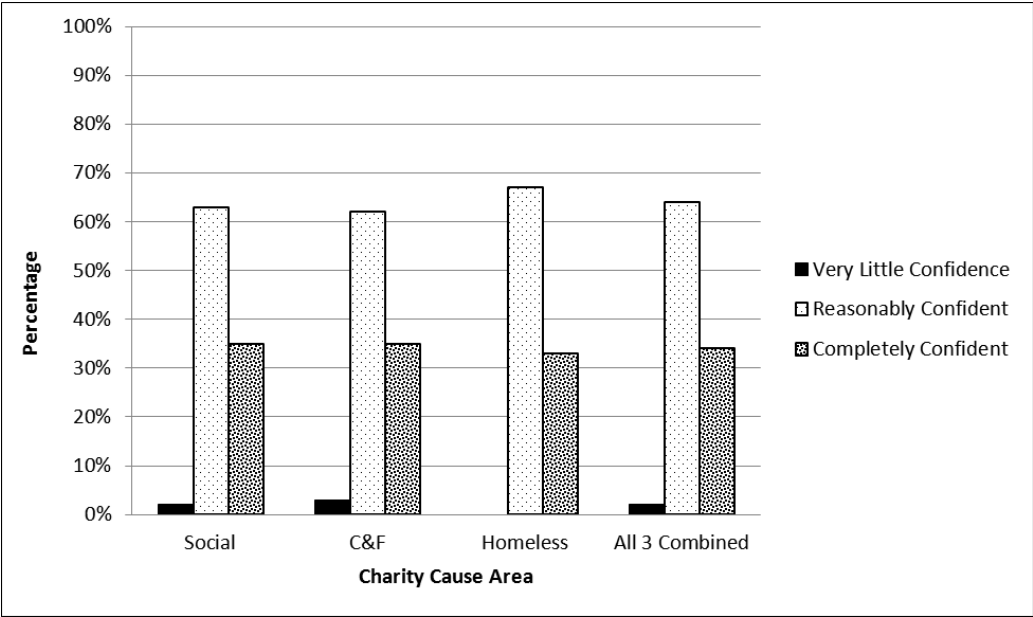
Quantitative: For Test One’s binary question, students chose “Yes” more than 90 percent of the time for all three cause areas combined and each area individually, as seen in Figure 4.1. Figure 4.2 shows that for the test’s Researcher Confidence Level section, “Reasonably Confident” was chosen the majority of the time, at over 60 percent, while “Completely Confident” was chosen less than 40 percent of the time, and “Very Little Confidence” less than 10 percent of the time. These findings held for all three cause areas combined and each area individually.

Figure 4.1: Percentage Breakdown of Element One – Test One: Is the Causal Logic Plausible?
Binary Question



n=174

Figure 4.2: Percentage Breakdown of Element One – Test One: Is the Causal Logic Plausible?
Researcher Confidence Level Responses

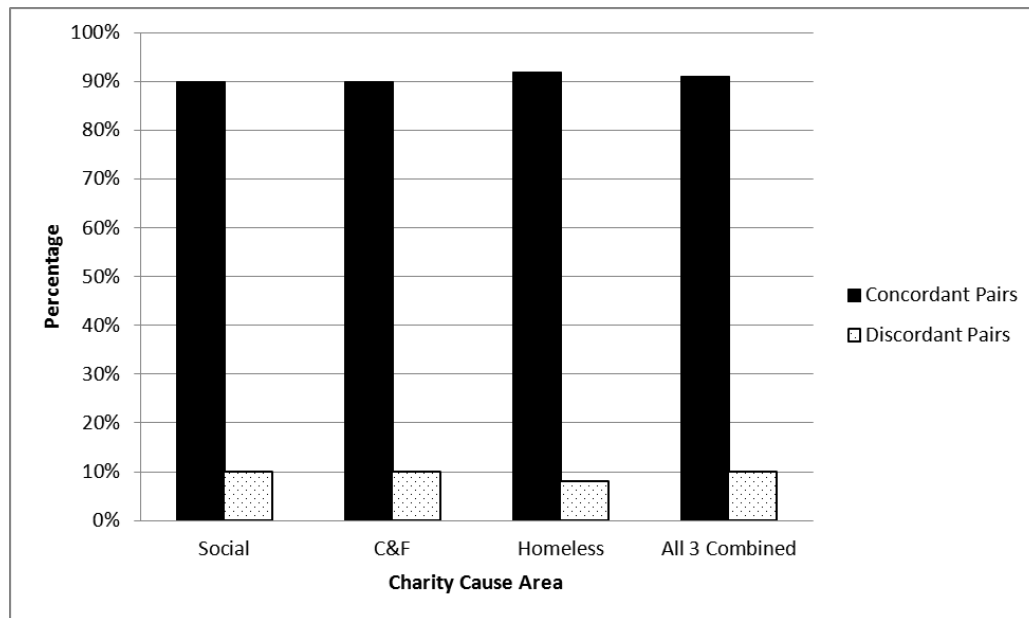


n=174

Qualitative: Students had concerns related to subjectivity for Test One. First, there was confusion regarding whether or not the test should be applied to the organization as a whole or to its individual programs. Second, students felt the way someone interprets “causal logic” and whether it is “plausible” will vary among raters. Although the vast majority of students answered “Yes” to the binary question, “Is the causal logic plausible?” some students reported that using a binary rating system was challenging, as they were not sure if an organization had “enough” of this element to receive the “Yes” answer. For Test One, students chose “Reasonably Confident” the majority of the time, both for all cause areas combined and for each area individually, which might reflect these concerns.

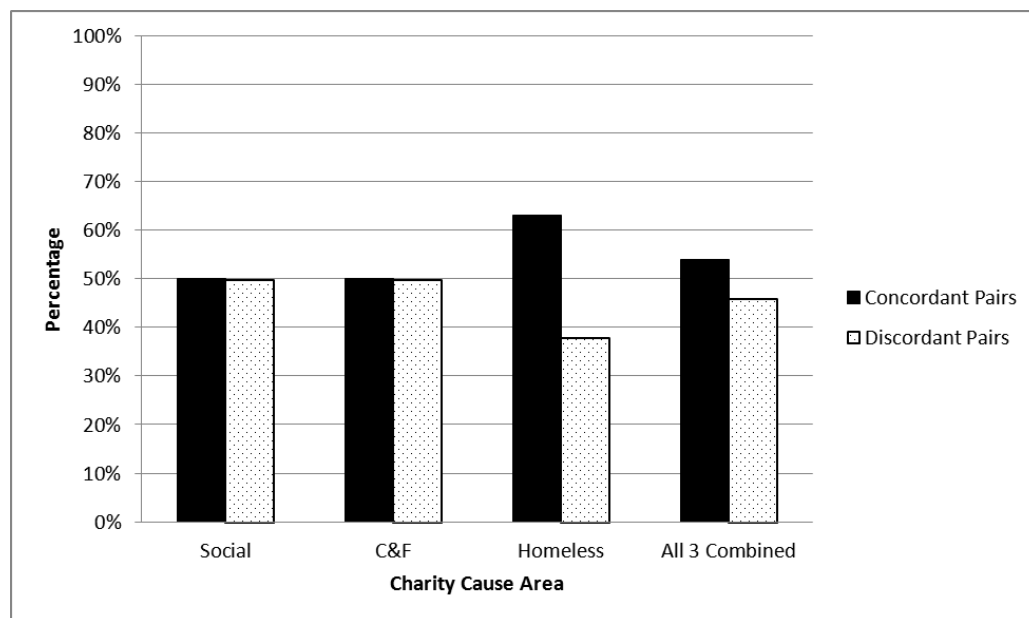
Inter-Rater Reliability: Figure 4.3 shows that, for all cause areas combined and each area individually, the vast majority of student pairs answered the same way for Test One’s binary question. According to Figure 4.4, however, the percentage breakdown between concordant and discordant pairs was more evenly split for the Researcher Confidence Level question. This means there was a great deal of variability in how confident students felt about their answers to this test. There was slightly less variability in students’ Confidence Level responses for Homeless Services organizations, though, as more than 60 percent of the responses were concordant, as Figure 4.4 reflects. Figure 4.5 shows that when there was a difference between how two students answered the Confidence Level question for a charity, the average difference equaled one rating level for all three cause areas combined and each area individually. This means that, even though there were many situations in which two students had different answers to the Confidence Level question, the differences between answers amounted to a small number (one level) rather than as large as two or three.

Figure 4.3: Percentage Breakdown of Inter-Rater Reliability for Element One – Test One: Is the Causal Logic Plausible? Binary Question Responses



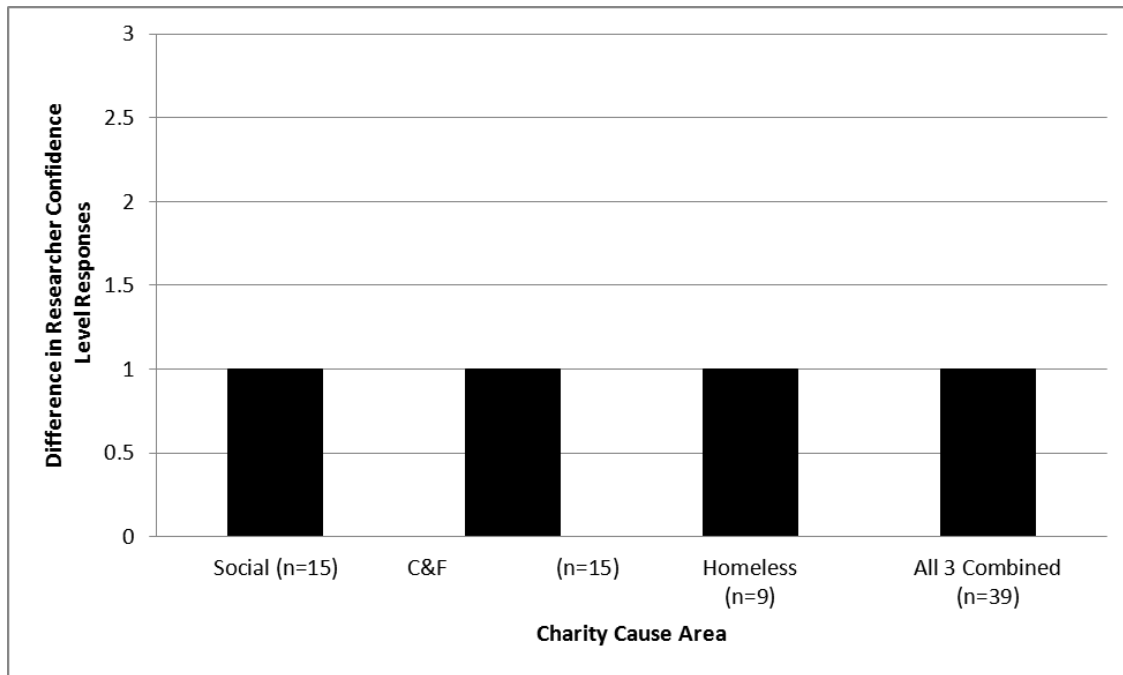
n=84

Figure 4.4: Percentage Breakdown of Inter-Rater Reliability for Element One – Test One: Is the Causal Logic Plausible? Researcher Confidence Level Responses



n=84

Figure 4.5: Average Difference in Researcher Confidence Level Responses for Discordant Pairs in Element One – Test One: Is the Causal Logic Plausible? (On a scale from 1 to 3)



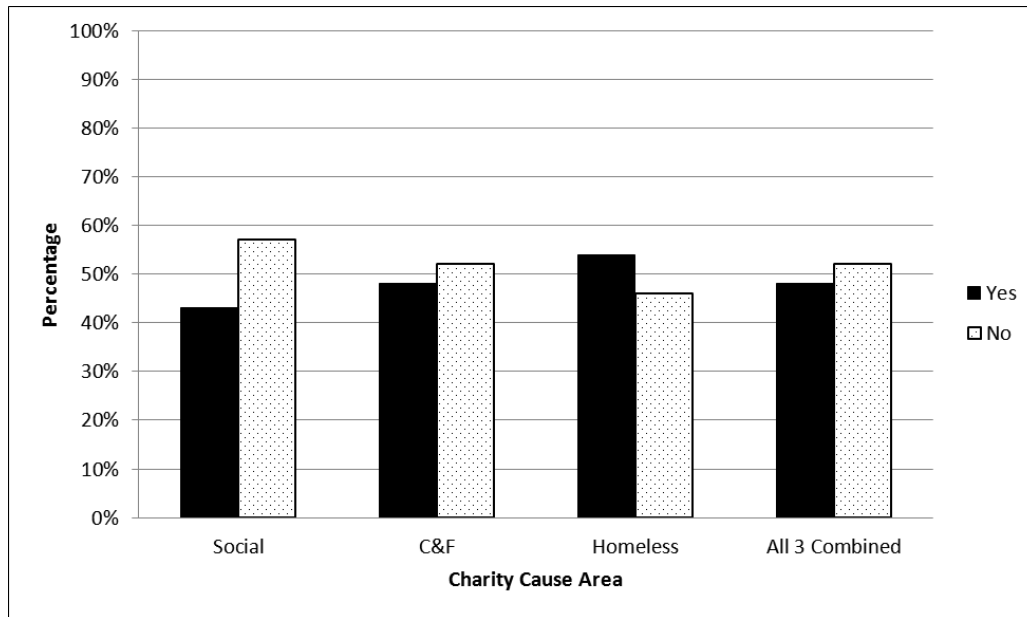
n provided = Total # of Discordant Pairs

Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)?

Quantitative: For Test Two’s binary question, student responses were almost evenly split between “Yes” and “No,” as seen in Figure 4.6. Students selected “No” for slightly more than 50 percent of the charities for all cause areas combined and for Children’s and Family Services charities, while it was higher for Social Services charities, at a little over 60 percent. Only Homeless Services charities had more “Yes” than “No” responses, but not by a large margin, at 54 percent to 46 percent. Figure 4.7 shows that for all three cause areas combined, “Reasonably Confident” was chosen the majority of the time, at over 65 percent, while “Completely Confident” was chosen less than 30 percent of the time, and “Very Little Confidence” less than 10 percent of the time. Although the percentage distribution among the

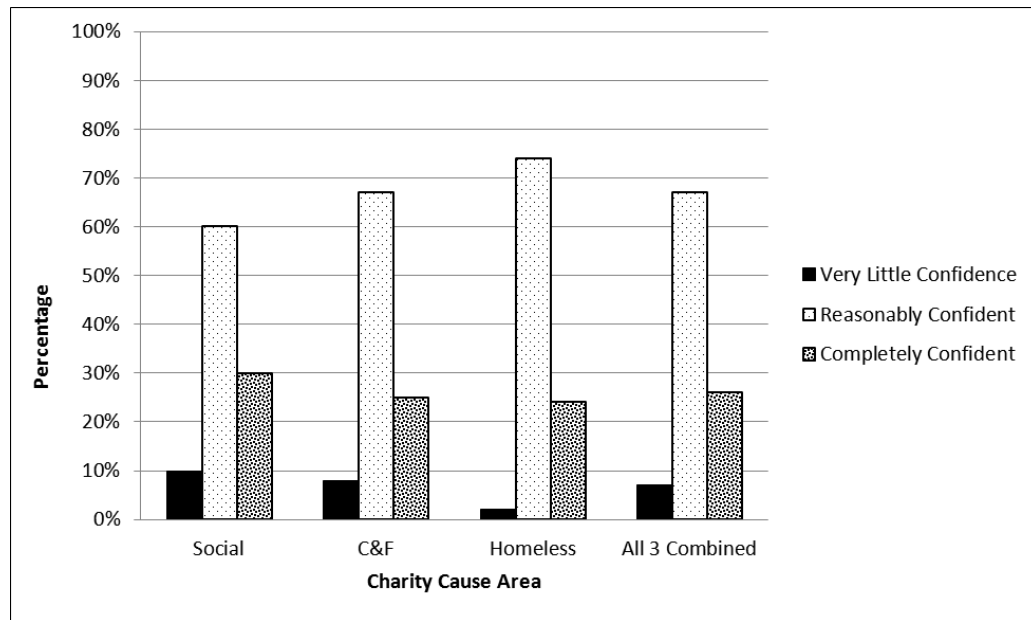
three confidence level answers varied slightly for each cause area individually, “Reasonably Confident” was still chosen the majority of the time, followed by “Completely Confident,” and then “Very Little Confidence,” as Figure 4.7 reflects.

Figure 4.6: Percentage Breakdown of Element One – Test Two: Is There Some Indication of How Much of the Action is required to Produce What Effects (i.e., dosage)? Binary Question



n=174

Figure 4.7: Percentage Breakdown of Element One – Test Two: Is There Some Indication of How Much of the Action is required to Produce What Effects (i.e., dosage)? Researcher Confidence Level Responses



n=174

Qualitative: Several students noted they could not find any information regarding dosage, which may have contributed to the fact that the distribution between “Yes” and “No” was almost evenly split for all three cause areas combined and each area individually. Students’ concerns in Test Two related to subjectivity and the binary rating system. Raters sometimes have to make inferences in order to answer questions, which contributes to the subjectivity issue. Students had difficulty with Test Two because, although the dosage is obvious in many situations, many charities do not explicitly state what the dosage is or provide a time frame. Students were not sure if they were expected to answer “No” when websites lacked an explicit and specific statement related to dosage. Students who are familiar with nonprofit organizations or with certain types of charities may be able to discern the dosage when it is not plainly stated, while others who are not as familiar may not be able to do so. Some students

noted how a person rates an organization depends largely on what part of the website he/she examines. Some students attributed low confidence levels to the fact that organizations' websites contained large quantities of information, which made them unsure about how other people would answer the same questions. As with Test One, students were not sure about whether or not to apply the test to the organization as a whole or to its individual programs.

Organizations with multiple programs may provide dosage information for some programs and not others, and students were not sure how to answer the binary question in these cases. Other students had difficulty answering Test Two because of the nature of some charities' activities. Quantitative dosage amounts are not – and perhaps cannot be -- described and students felt that having to answer "No" was unfair to these types of charities. For example, the mission of the charity SAGE Eldercare is to "support the independence, well-being and quality of life of older adults, their families and caregivers, through the provision of client-centered health, social and support services" (SAGE Eldercare, p. 2). After looking through the charity's website without finding explicit information regarding dosage, the student answered "No" for the binary question. She then wrote in the 'Main Challenge to Confidence in Answers' open-ended section that she did not know how the organization could quantify a dosage: "A 'dosage' is not explicit. But I am not sure this is entirely clear. How would this group say 'it will take this many at-home visits to increase the quality of life' and so forth? Based on some of their terminology and information, some may answer this as 'Yes'" (from Charity Navigator 3.0 Tool results, included in raw data).

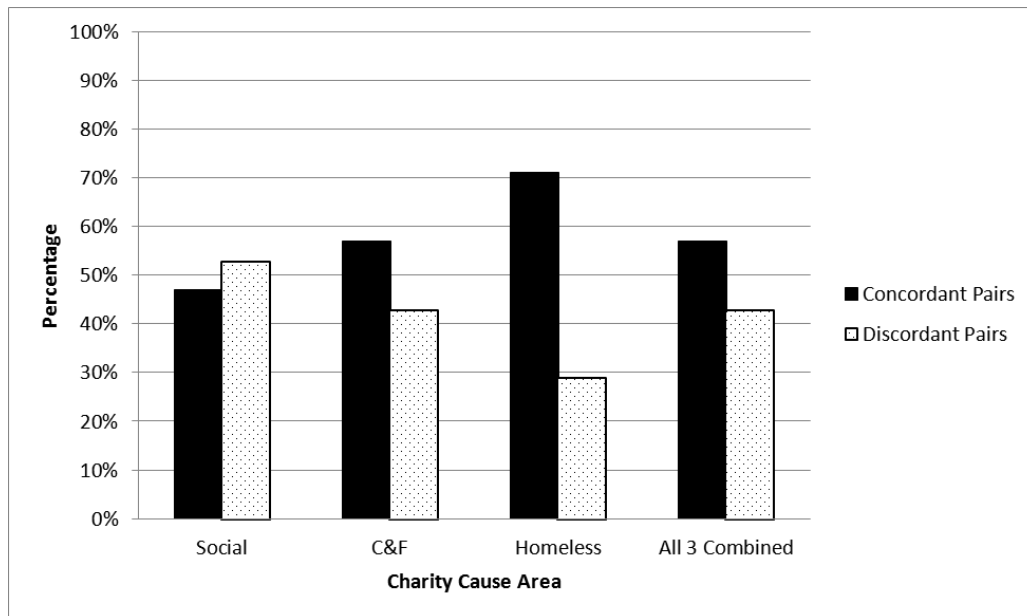
For all three cause areas combined, students' answers were approximately evenly distributed between "Yes" and "No," which illustrates that these concerns were especially

prevalent for questions related to dosage. There was some variation across the three charity cause areas for the binary question, as the percentage distribution of responses between “Yes” and “No” was nearly 50/50 for Children’s and Family Services and Homeless Services, while the percentage distribution for Social Services was almost 40/60. This variation could be due to actual differences in how these different types of charities conduct results-reporting, or to students’ difficulty in applying the Tool’s questions. Subjectivity and binary rating system concerns may have contributed to the fact that for Test Two, students said they were only “Reasonably Confident” the majority of the time, both for all three cause areas combined and each area individually.

Inter-Rater Reliability: Figure 4.8 shows that for all three cause areas combined, there were more concordant student pairs than discordant pairs for Test Two’s binary question. The percentage distribution between concordant and discordant pairs for all cause areas combined and Children’s and Family Services were both close to 50/50, while for Homeless Services the percentage distribution was closer to 70/30. More student pairs had different answers than identical answers for Social Services Organizations, although the percentage distribution was very close to 50/50, as Figure 4.8 reflects. For the Researcher Confidence Level question, there were more discordant than concordant pairs for all cause areas combined, but for Homeless Services Organizations, there were more concordant than discordant pairs, as shown in Figure 4.9. However, for Homeless Services Organizations, the percentage distribution between concordant and discordant responses was relatively close to 50/50. This means there was a great deal of variability in how confident students felt about their answers to this test and the extent of variation was not consistent across the three charity cause areas.

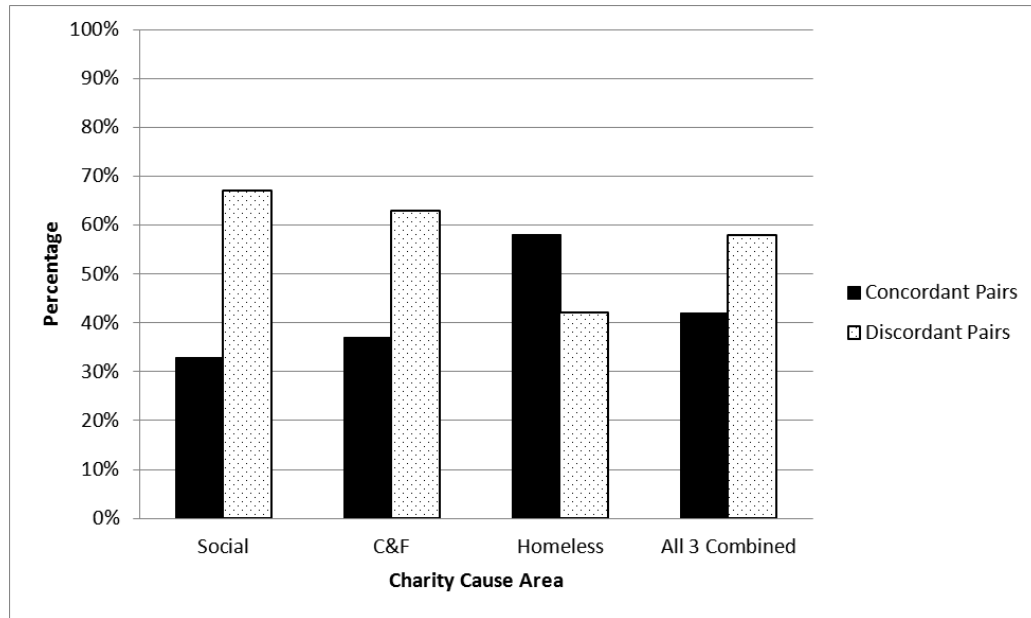
Just as with Test One, when there was a difference between how two students answered the Confidence Level question for a charity, the average difference equaled one level, for all three cause areas combined and each area individually, as seen in Figure 4.10. Thus, even though there were many discordant pairs for the Confidence Level question, the differences among the answers amounted to a small number (one level) as compared to what they could have been (two or three).

Figure 4.8: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Two: Is There Some Indication of How Much of the Action is required to Produce What Effects (i.e., dosage)? Binary Question Responses



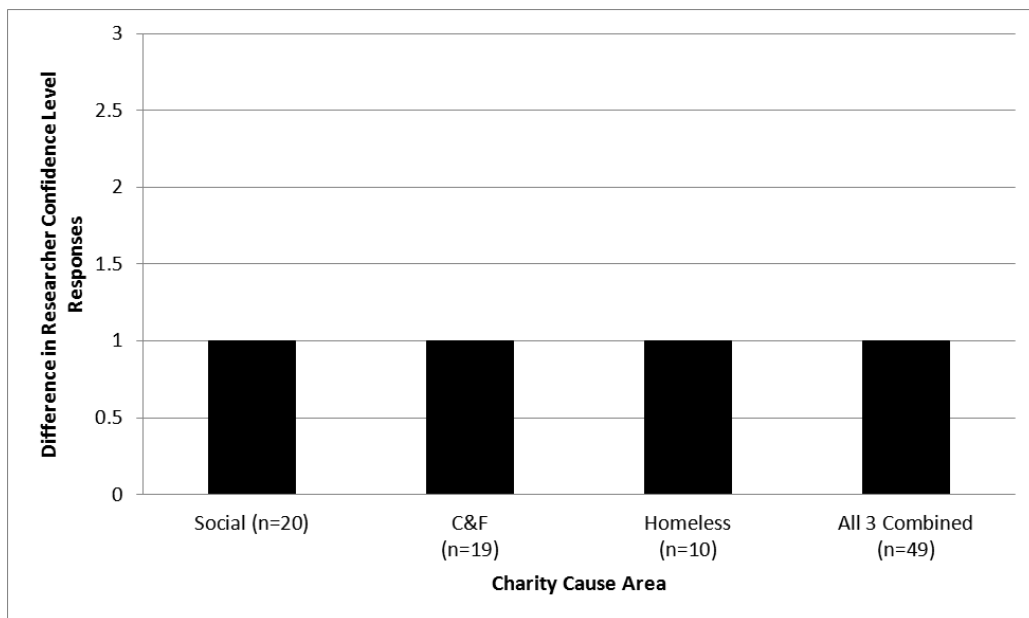
n=84

Figure 4.9: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Two: Is There Some Indication of How Much of the Action is required to Produce What Effects (i.e., dosage)? Researcher Confidence Level Responses



n=84

Figure 4.10: Average Difference in Researcher Confidence Level Responses for Discordant Pairs in Element One – Test Two: Is There Some Indication of How Much of the Action is required to Produce What Effects (i.e., dosage)? (On a scale from 1 to 3)

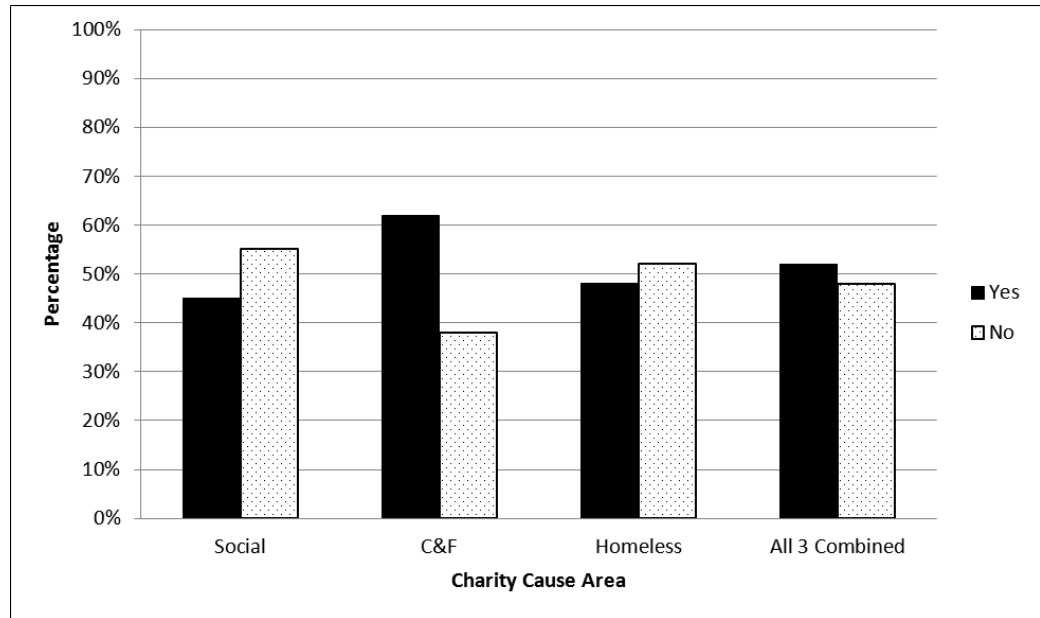


n provided = Total # of Discordant Pairs

Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)?

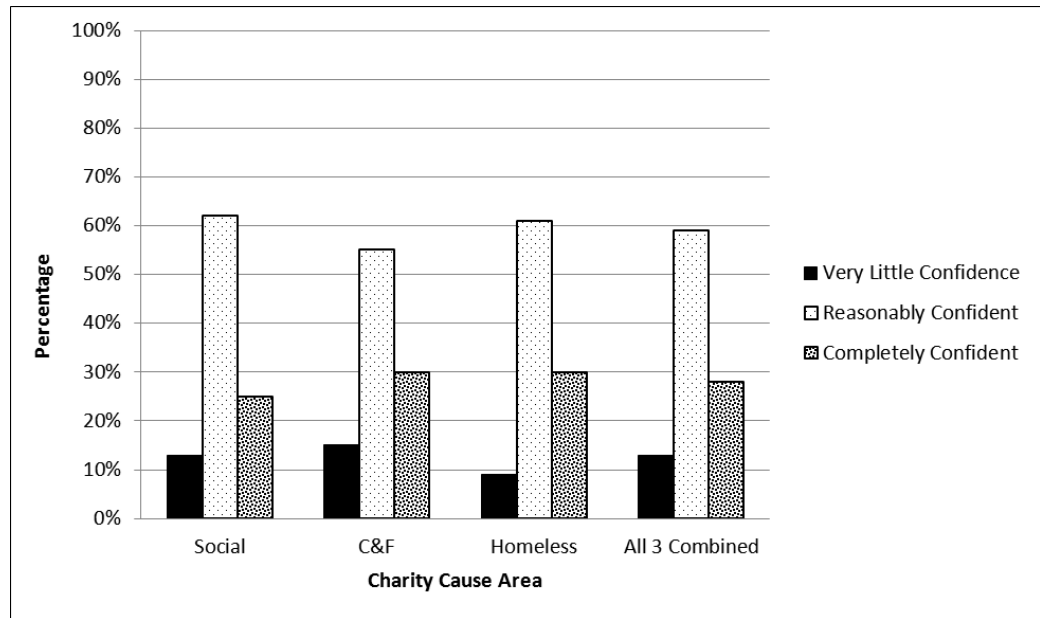
Quantitative: Figure 4.11 shows that, for Test Three's binary question, student responses for Social Services organizations, Homeless Services organizations, and all three cause areas combined were almost evenly split between "Yes" and "No," with only all cause areas combined receiving more "Yes" than "No" responses. Children's and Family Services organizations was the only cause that had more "Yes" than "No" responses, with the Yes/No percentage distribution slightly higher than 60/40. Figure 4.12 shows that for all three cause areas combined in the Researcher Confidence Level section, "Reasonably Confident" was chosen the majority of the time, at over 55 percent, while "Completely Confident" was chosen less than 30 percent of the time, and "Very Little Confidence" less than 15 percent of the time. Just as with Test Two, although the percentage distribution between the three confidence level answers varied slightly for each cause area, students still chose "Reasonably Confident" the most often, followed by "Completely Confident," and then "Very Little Confidence," as Figure 4.12 illustrates.

Figure 4.11: Percentage Breakdown of Element One – Test Three: Is There Some Indication that the Logical Inference in (1) is based on Reasonable Evidence (i.e., references to other programs that have demonstrated success with this logic)? Binary Question



n=174

Figure 4.12: Percentage Breakdown of Element One – Test Three: Is There Some Indication that the Logical Inference in (1) is based on Reasonable Evidence (i.e., references to other programs that have demonstrated success with this logic)? Researcher Confidence Level Responses



n=174

Qualitative: Student subjectivity concerns in Test Three were related in part to terminology, specifically “reasonable evidence,” which some students felt might be interpreted differently depending on the rater. As with Test Two, students also felt that raters may have to make inferences in order to answer this test’s questions, as organizations commonly provide information about their history, but do not openly assert that this is evidence of causal logic.

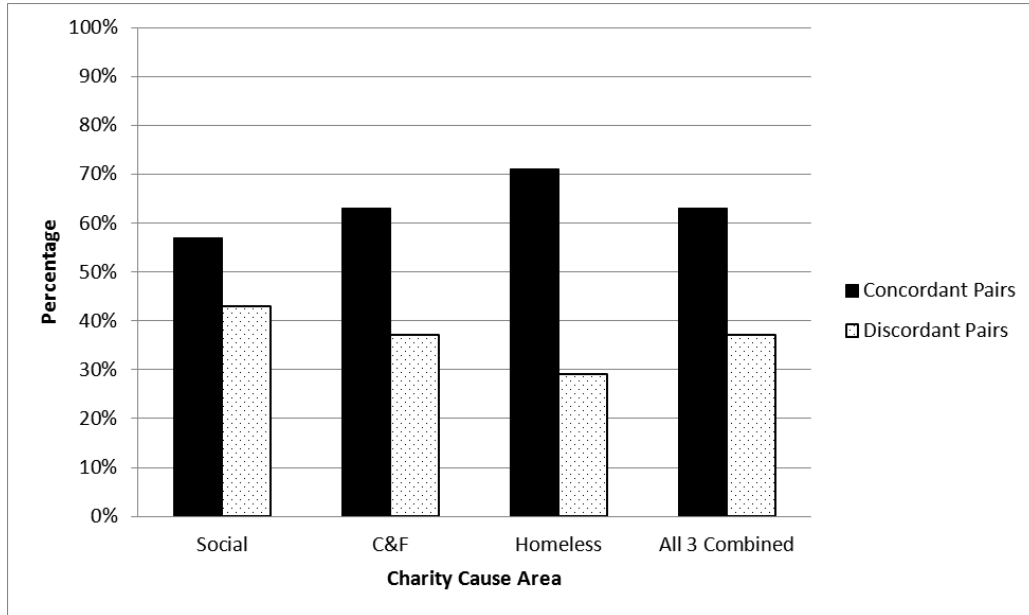
There were also concerns related to the binary rating system. Quite a few students said they felt that the example provided within the Tool for this test—“references to other programs that have demonstrated success with this logic”—may lead to many “No” answers because finding this kind of evidence on websites is rare (or in fact, the amount of research in some fields has been small or inconclusive). One reason for the variation among responses was that some students, with knowledge of the nonprofit sector, knew to look for other signs of this element before simply choosing “No.” As with Test Two, the confusion may be partly responsible for the fact that responses were approximately evenly distributed between “Yes” and “No.” There was some variation across the three charity cause areas, as the percentage distribution was nearly 50/50 for Social Services and Homeless Services, while the percentage distribution for Children’s and Family Services was approximately 60/40. Just as with Test Two, this variation could be due to actual differences in how these different types of charities conduct results reporting, but they may also be due to students’ uncertainties about this test. Just as in previous tests, subjectivity and binary rating system concerns may have been responsible for the fact that for all three cause areas combined and each area individually, “Reasonably Confident” was chosen more frequently than the other two confidence level answers in Test Three.

Inter-Rater Reliability: Figure 4.13 shows that for all three cause areas combined and each area individually, there were more concordant student pairs than discordant pairs for Test Three's binary question. However, the percentage distribution between concordant and discordant pairs varied by charity cause area. The highest amount of inconsistencies in student pairs' responses was observed for Social Services organizations, followed by all three cause areas combined and Children's and Family Services organizations, while the smallest number of inconsistencies was seen in student responses for Homeless Services organizations, as Figure 4.13 reflects. For the Researcher Confidence Level question, there were more discordant than concordant pairs for all three cause areas combined, Children's and Family Services organizations, and Homeless Services organizations, while there were more concordant than discordant pairs for Social Services organizations, as seen in Figure 4.14. This means there was a great deal of variability in how confident students felt about their answers to this test and that the extent of variation was not consistent across the three charity cause areas.

Just as with Tests One and Two, when there was a difference between how two students answered the Confidence Level question for a charity, the average difference equaled one level for all three cause areas combined and each individually, as seen in Figure 4.15. Thus, just as with the previous two tests, though there were many discordant pairs for the Confidence Level question, the differences between the answers amounted to a small number (one level) as compared to what they could have been (two or three).

Figure 4.13: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Three: Is There Some Indication that the Logical Inference in (1) is based on Reasonable Evidence (i.e., references to other programs that have demonstrated success with this logic)?

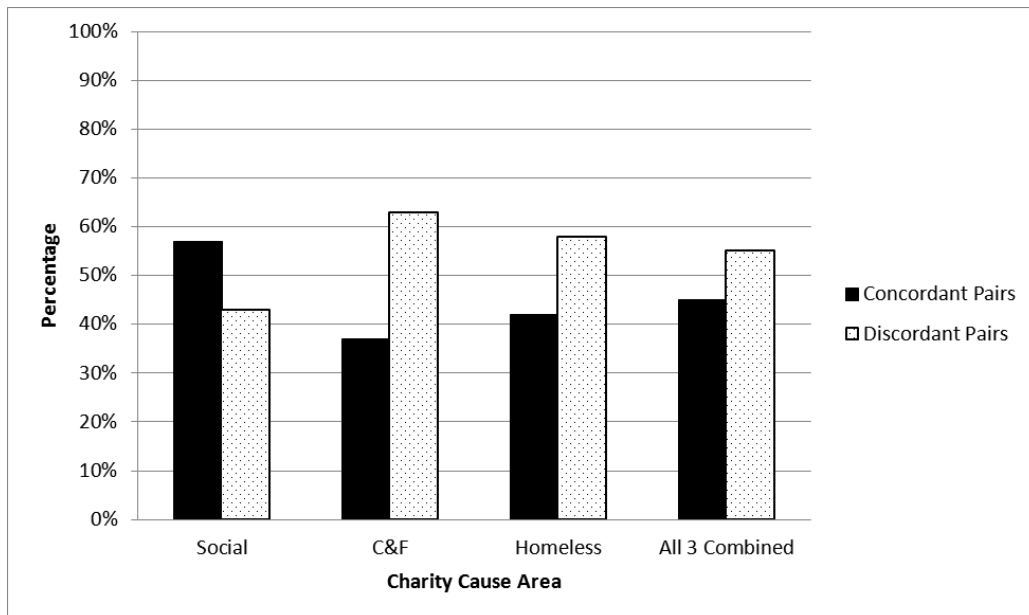
Binary Question Responses



n=84

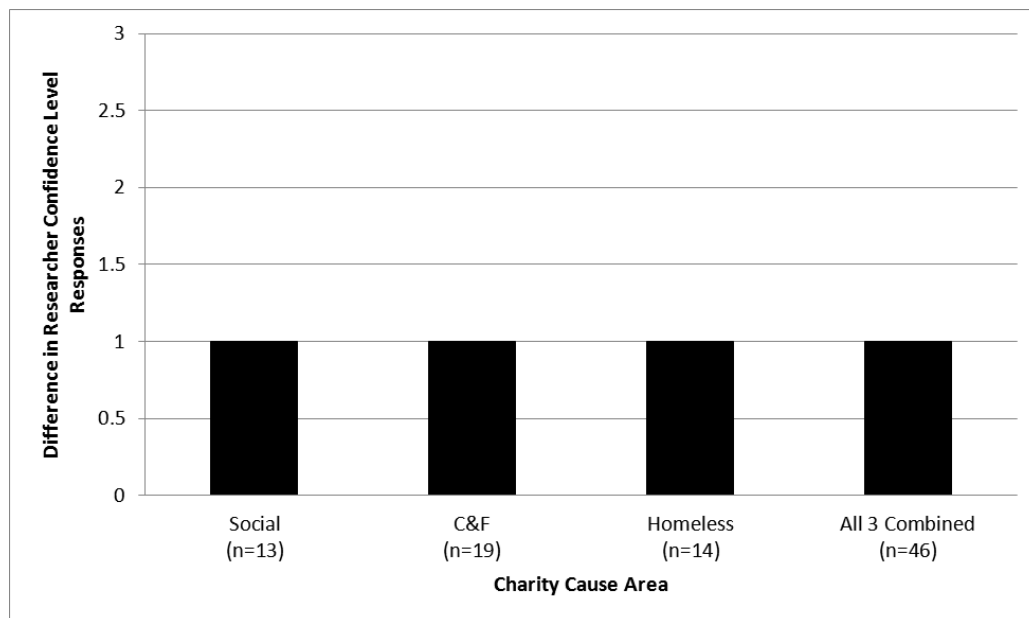
Figure 4.14: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Three: Is There Some Indication that the Logical Inference in (1) is based on Reasonable Evidence (i.e., references to other programs that have demonstrated success with this logic)?

Researcher Confidence Level Responses



n=84

Figure 4.15: Average Difference in Researcher Confidence Level Responses for Discordant Pairs in Element One – Test Three: Is There Some Indication that the Logical Inference in (1) is based on Reasonable Evidence (i.e., references to other programs that have demonstrated success with this logic)? (On a scale from 1 to 3)



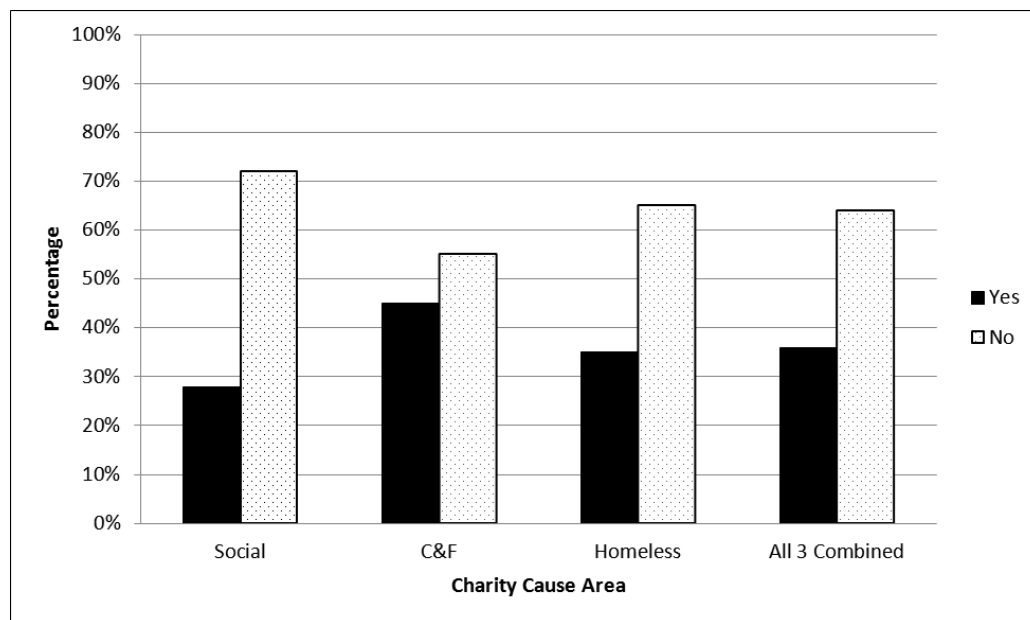
n provided = Total # of Discordant Pairs

Test Four: Are there specified measures (indicators) to be collected and a plan to do so?

Quantitative: In contrast to the other three tests, “No” was chosen more frequently than “Yes” in Test Four’s binary question, both for all three cause areas combined and each cause area individually, as seen in Figure 4.16. Also, only responses for Children’s and Family Services organizations were almost evenly split between “Yes” and “No,” as the Yes/No percentage distribution was approximately 30/70 for Social Services organizations, 35/65 for Homeless Services organizations, and close to 35/65 for all cause areas combined. Figure 4.17 shows that for all three cause areas combined in the test’s Researcher Confidence Level section, “Reasonably Confident” was chosen the majority of the time, at over 50 percent, while

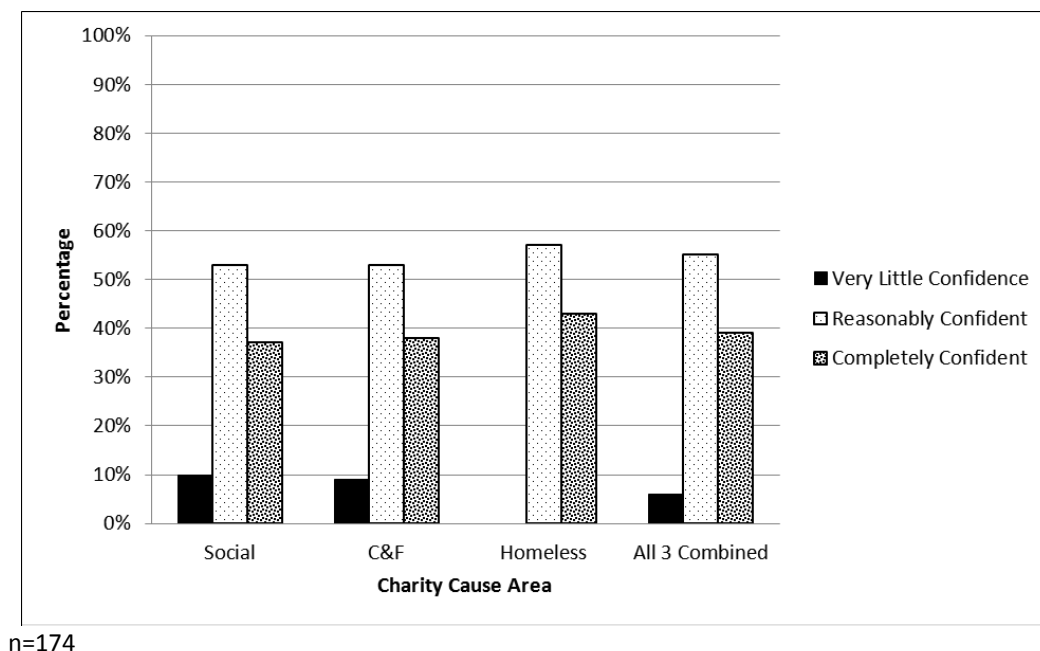
“Completely Confident” was chosen less than 40 percent of the time, and “Very Little Confidence” less than six percent of the time. Just as with Tests Two and Three, although the percentage distribution between the three confidence level answers varied slightly for each cause area individually, students still chose “Reasonably Confident” the most often, followed by “Completely Confident,” and then “Very Little Confidence,” as seen in Figure 4.17. The Confidence Level findings for this Test were atypical, however, as the gap between the percentages of students who chose the “Reasonably Confident” and “Completely Confident” answer choices for all three cause areas combined and each area individually was smaller than it was in any of the other three tests.

Figure 4.16: Percentage Breakdown of Element One – Test Four: Are There Specified Measures (indicators) to Be Collected and a Plan to Do So? Binary Question



n=174

Figure 4.17: Percentage Breakdown of Element One – Test Four: Are There Specified Measures (indicators) to Be Collected and a Plan to Do So? Researcher Confidence Level Responses



Qualitative: Students felt that subjectivity was an issue for Test Four for several reasons. First, students were not sure how to answer the test when organizations provided “specified measures to be collected and a plan to do so” for some of their programs, but not for others. Also, because the charities’ websites are not standardized, the information that raters are able to collect will vary depending on individuals’ research skills, prior knowledge, and experience, which will impact the ratings. Other students noted that large organizations often have a considerable number of reports and publications available on their websites that can provide useful information for this test. However, because reading every single report or publication is time-consuming, and there is no mechanism to ensure that all raters are examining the same documents, there is a threat to the reliability of the ratings. Concerns about having to make inferences were due to the fact that, although students often found

statistics and numbers about charities' services, organizations rarely explicitly stated that they used formal processes and had a plan in place to gather data. Students said it was clear that organizations are collecting data, but students could not find evidence of official policies and procedures regarding the issue.

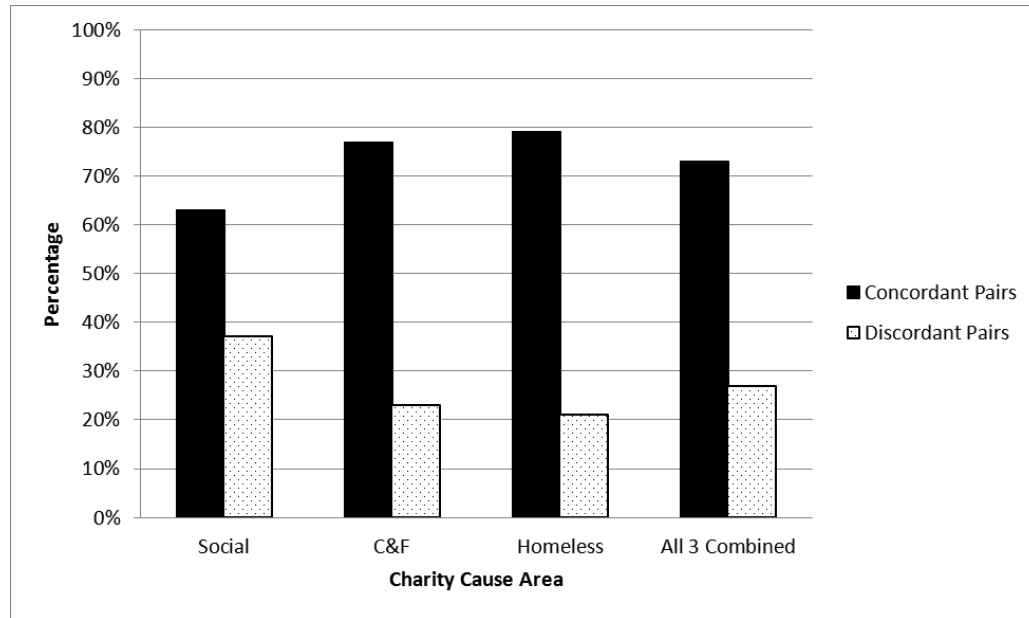
Students also said they felt the wording of this test made using a binary answering system difficult because the test was actually asking two different questions: are there specified measures (indicators) to be collected, and is there a plan to do so? Students were unsure about how to respond when they found evidence to answer one of the questions but not another. The fact that "No" was chosen 64 percent of the time for Test Four might reflect students' uncertainty about the question, or it could also suggest that Charity Navigator may not be able to expect the average charity to have both specified measures (indicators) to be collected and a plan to do so included on a website. Although "No" was chosen the majority of the time for all three cause areas for Test Four, there was some variation between the different types of organizations. The percentage distribution of Yes/No for Children's and Family Services was close to 50/50, while for Social Services it was approximately 30/70, and for Homeless Services was nearly 35/65. Again, this variation could be due to actual differences in how these different types of charities conduct results-reporting, or to students' confusion about how to answer this test. These issues undoubtedly contributed to the fact that for Test Four, students reported being only "Reasonably Confident" the majority of the time, for all three cause areas combined and each area individually, which is a finding consistent with the previous tests.

Inter-Rater Reliability: Figure 4.18 shows for all three cause areas combined and each area individually, there were more concordant student pairs than discordant pairs for Test

Four's binary question. However, the percentage distribution between concordant and discordant pairs varied by charity cause area. Social Services responses had the highest percentage of discordant pairs, followed by all three cause areas combined. For the Researcher Confidence Level question, there were more discordant than concordant pairs for all three cause areas combined and each area individually, as seen in Figure 4.19. The percentage distribution between concordant and discordant pairs for all three cause areas combined, Social Services organizations, and Children's and Family Services organizations were all somewhat close to 50/50, while for Homeless Services organizations the percentage of responses that were discordant was significantly larger than the percentage of responses that were concordant, as Figure 4.19 reflects. This means there was a great deal of variability in how confident students felt about their answers to this test and that the extent of the variation was not consistent across the three charity cause areas.

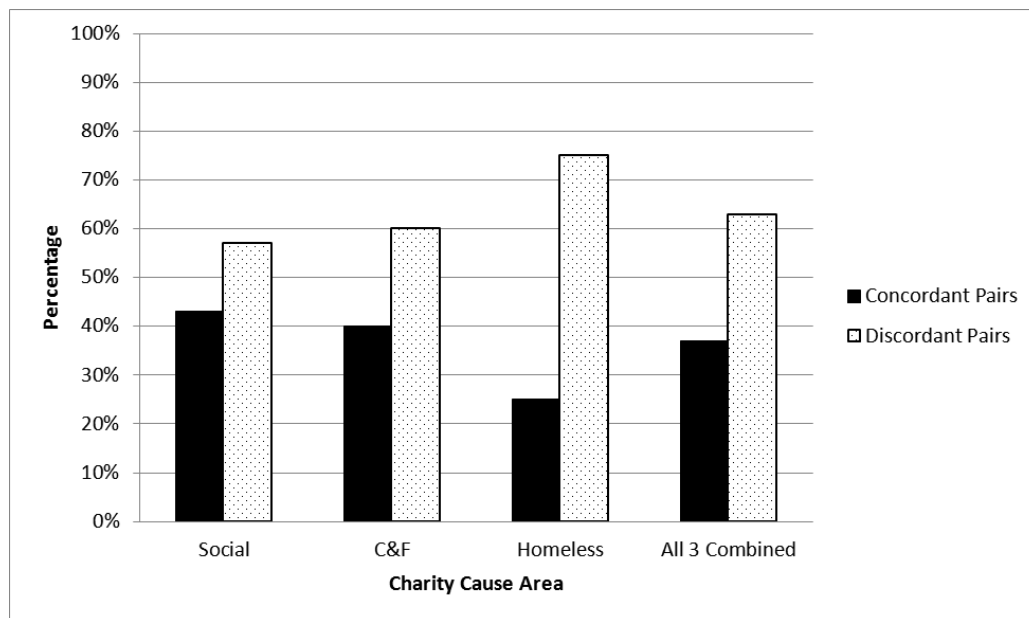
In keeping with all three previous tests' findings, when there was a difference between how two students answered the Confidence Level question for a charity, the average difference equaled one level both for all three cause areas combined and each individually, as seen in Figure 4.20. Thus, even though there were more discordant than concordant pairs for the Confidence Level question in Test Four, the differences between the answers in the discordant pairs amounted to a small number (one level) as compared to what it could have been (two or three).

Figure 4.18: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Four: Are There Specified Measures (indicators) to Be Collected and a Plan to Do So? Binary Question Responses



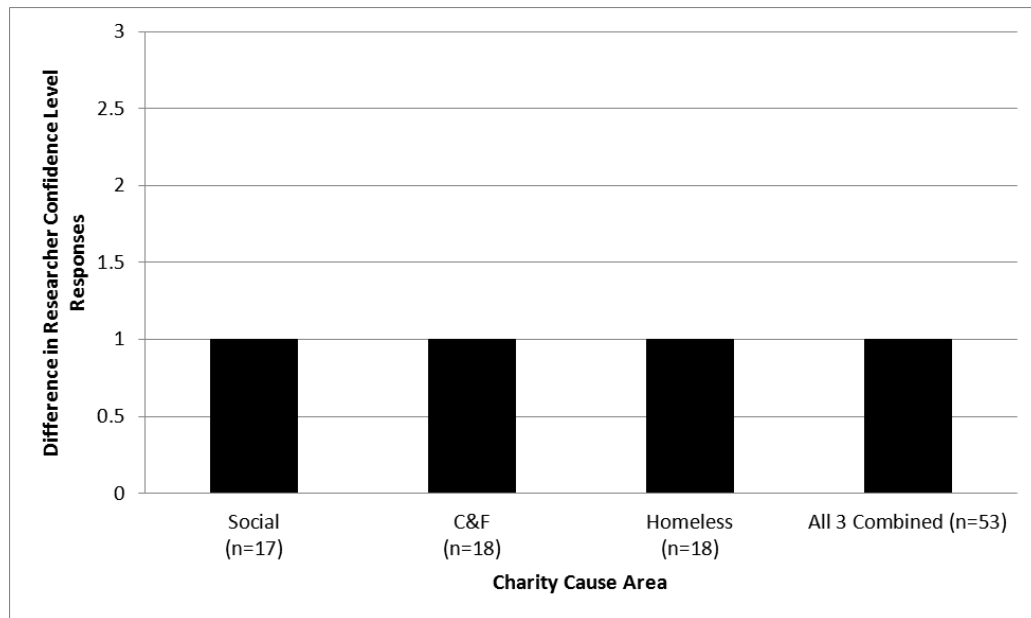
n=84

Figure 4.19: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Four: Are There Specified Measures (indicators) to Be Collected and a Plan to Do So? Researcher Confidence Level Responses



n=84

Figure 4.20: Average Difference in Researcher Confidence Level Responses for Discordant Pairs in Element One – Test Four: Are There Specified Measures (indicators) to Be Collected and a Plan to Do So? (On a scale from 1 to 3)



n provided = Total # of Discordant Pairs

Observations from the Background Research Tool.

As stated in the methodology section, the Capstone Group initially created the BRT to understand the state of publicly available results-reporting information from nonprofits. The BRT was not intended to serve as an alternative tool, nor were the questions developed in comparison to the CNRT. Rather the BRT was meant to add to the range of items to look for on a charity's website. (For more detailed analysis, see the Background Research Tool Frequency Table related to Element One in Appendix C.)

Students collected results reporting information using the Background Research Tool for 89 of the charities identified by CN, as shown in Figure 4.21. (While the initial assigned sample size was 90, one charity's website was nonoperational.) Sample sizes for each of the assigned

cause areas are as follows: Children's and Family Services- 29, Social Services - 30, Homeless Services - 30.

Mission: The first type of additional information related to Element One that students were asked to search for and review was each charity's mission. Students were asked to "Consider if the mission is related to impact, goals." Foundations and information intermediaries regard effective charities as having clearly defined mission statements that are aligned with organizational goals (Miller, 2012). Henry Migliore refers to the mission statement as a "reason for being" and the communication tool for who the charity is and what the charity intends to do (Migliore, et al., 1995).

Of the 89 charities that were researched, 96 percent (85 out of 89) published a mission statement. However, it was not always clear how the stated mission was related to the charity's impact or goals, especially if it was vague or not specific to programs listed. Of these charities, 28 came from Social Services, 29 were Children's and Family Services, and 32 were Homeless Services.

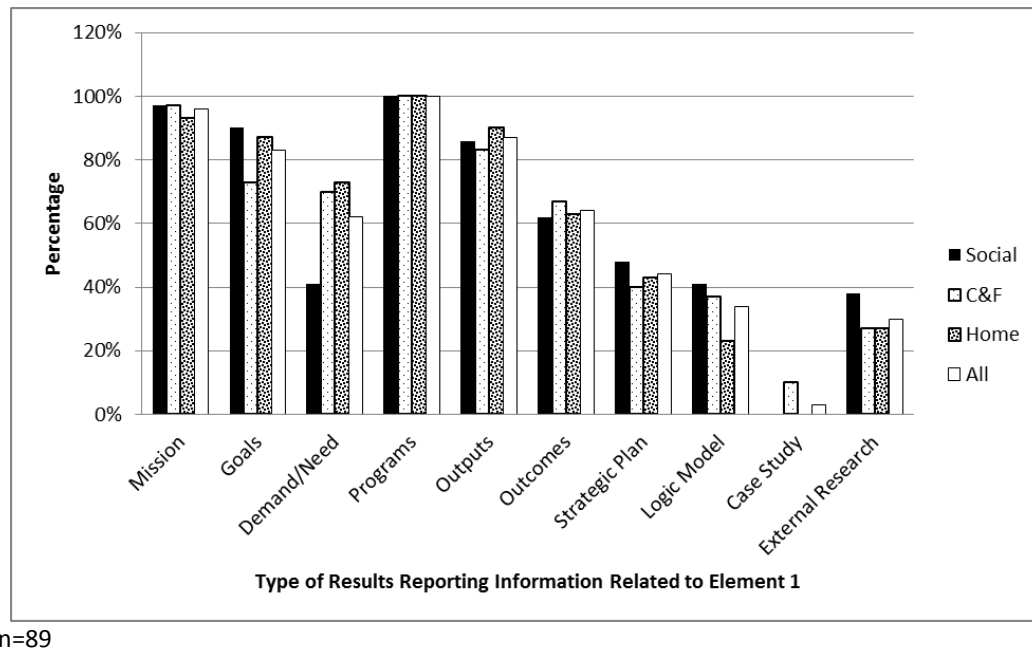
Goals: The second type of information students were asked to search for and review was the charity's goals. The Capstone Group defined goals as published, short-term steps to accomplish a mission. Once a charity clearly communicates its purpose through a well-crafted mission statement, it is able to define intended outcomes and specific program goals. Students were asked to consider "Are intended outcomes described?" Short-term and intermediate goals are considered benchmarks that could lead to long-term impact.

Of the 89 charities that were researched, 83 percent (74 out of 89) published goals. Of these charities, 26 came from Social Services, 22 were Children's and Family Services, and 26 were Homeless Services.

Strategic Plan: The third type of information students were asked to search for and review was the charity's strategic plan. Students were prompted to look for any evidence of organizational planning through the following guiding question: "Even if there isn't a 'strategic plan,' is there evidence of organizational planning at all?" Five of the seven information intermediaries indicated communicating charities' various strategies was important because it allows for a high degree of transparency about the approach a charity intends to employ in accomplishing its long-term goals, which, in turn, should strengthen a charity's credibility. Strategic planning also helps define a charity's capacity and how it intends to grow, and reflects a charity's commitment to reassessing its operations. Strategies can be incorporated within various types of plans, in a variety of different ways, and for diverse reasons, depending on the needs of the organization (M. Allison and J. Kaye, 1997). In light of this, students were asked to look for any form of organizational planning.

Of the 89 charities that were researched, 44 percent (39 out of 89) published strategic plan and/or displayed evidence of organizational planning. Of these charities, 14 came from Social Services, 12 were Children's and Family Services, and 13 were Homeless Services.

Figure 4.21: Background Research Tool Analysis – Element One



Reference to Data/Research: The fourth type of published information students were asked to assess was whether the charity referenced any overarching data, research, or literature that provided justification for the programs and services it provides. Students were given the following guiding questions: “Does the charity compare its impact or products/services to data sources? Does it relate the programs/services it provides to literature that proves the impact of these programs/services?” Information intermediaries, like Standards for Excellence Institute, ask that charities re-evaluate the need for their organization every three to five years by revisiting its mission, as well as reflecting on societal changes and whether related programs should be developed or discontinued (Standards for Excellence Institute, 2012). However, none of the intermediaries directly asks for a charity to compare its programs/services to data, research literature, or other sources that might indicate if its programs/services have proven impact. Experts in the nonprofit field, like Migliore, highlight

the initial and ongoing importance of a charity's ability to clearly articulate the need for its programs/services and ultimately, its reason for existence (Migliore, 1995).

Of the 89 charities that were researched, 30 percent (27 out of 89) referenced overarching data, research, or other kinds of literature on their websites. Of these charities, 11 came from Social Services, eight were Children's and Family Services, and eight were Homeless Services.

Outputs: The fifth type of information that students were asked to search for and review was outputs. Students were given the following guiding questions: "Does the charity report how many clients were served? In an overall sense or does it break the figures down by each program service?" Intermediaries like InterAction include strands within their indicators, asking about the impact of outputs on program participants as a part of results measurement. Overall, outputs are an important indicator of how a charity links its inputs with intended outcomes through program activities, deliverables, and the number of participants. Students were asked to assess whether charities discuss outputs in terms of participants or in relation to various program services.

Of the 89 charities researched, 87 percent (77 out of 89) published information related to outputs on their websites. Of these charities, 25 came from Social Services, 25 were Children's and Family Services, and 27 were Homeless Services.

Outcomes: The sixth type of information that students were asked to search for and review was outcomes. Students were given the following guiding questions: "Does the charity report outcomes or overall impact at all? Does it differentiate between short-term, intermediate, and long-term outcomes? Are intended outcomes described?" GuideStar and

Charting Impact—as well as other information intermediaries—ask charities to go beyond reporting outputs to clearly communicate how outcomes contribute to long-term goals (Charting Impact, 2012). Identifying desired outcomes is considered an initial step to determining quality indicators for marking progress toward those outcomes. Indicators may vary across charities and cause areas; however, quality indicators would likely include ones that are specific, observable, relevant, time-bound, valid, and comprehensible (Urban Institute, 2012). Specific and comprehensible criteria enable the development of clear, concise, and easily understood indicators. Observable and time-bound criteria create practical, yet measurable, indicators that occur within a specific time period. Finally, relevant and valid criteria allow for an “unbiased and verifiable” collection of data that are significant to the program type and appropriately measure the important dimensions of program (Urban Institute, 2012).

These criteria, or similar criteria for quality indicators, can be utilized for various types of intended outcomes. Of the seven information intermediaries that include some form of results measurement, five include a statement regarding the need for defined program outcomes. Outcomes enable charities to more clearly define where they are in relation to fulfilling their mission. As described in Chapter Two, the majority of intermediaries addressing results measurement request that charities have pre-defined outcomes. Outcomes are the building blocks for measuring mission effectiveness (Mitchell, 2010). Additionally, if charities have not defined intended outcomes, then other components of results measurement that CN and information intermediaries are asking for, such as tracking progress towards those outcomes, are irrelevant because there is no end goal in mind.

Of the 89 charities researched, 64 percent (57 out of 89) published information related to outcomes on their websites. Of these charities, 18 came from Social Services, 20 were Children's and Family Services, and 19 were Homeless Services.

Programs: The seventh type of information students were asked to search for and review concerned the charity's programmatic elements. Students were given the following guiding questions: "What services/products does the charity provide? Are these activities linked to the intended outcomes? Are they related back to the needs and/or mission?" Charities can offer programs and/or services, but their purpose needs to be clearly understood and mission-driven. Charities are in the "people-changing" business (Gronbjerg, 2001). Programs and services offered by charities are intended to change individual or group knowledge, behavior, attitude, mindset, condition, etc. Therefore, programs and services need to be clearly linked to the intended outcomes—whether they are participant-centered or community-centered, or a change in behavior or condition (Urban Institute, 2012).

Of the 89 charities that were researched, 100 percent published information related to programmatic elements on their websites.

Demand and Need: The eighth type of information that students were asked to research was whether the charity published data or information related to showing the demand or need for the charity's programs or services. Students were given the following guiding questions: "Does it show a trend of increased demand/unmet need? Does it mention a time period in any of these cases? Has it performed a needs assessment?" Sargeant and Shang (2010) and The Fundraising School (2011) discuss the importance of articulating a demand for services and the societal need being met, in part, to provide a rationale for donor support. Established, larger organizations may have a greater capacity to perform a needs assessment, but all charities should be able to describe the need that is being addressed.

Of the 89 charities that were researched, 62 percent (55 out of 89) published data or information related to showing the demand or need for the charity's services on their websites. Of these charities, 12 represented Social Services, 21 were Children's and Family Services, and 22 were Homeless Services.

Logic Models: The ninth type of published information that students were asked to search for and review was a logic model or illustration of the relationship between input, activities, outputs, and outcomes. Students were prompted by the following guiding question: "Consider attribution – does the charity make an effort to show how its services/products/programs directly relate to impact/outcomes?" This question addresses components related to output and outcome alignment. Although charities are not necessarily expected to portray logic models in order to obtain high ratings by information intermediaries and accreditation organizations, those organizations often ask for some description of the linkage between program activities and desired outcomes.

Of the 89 charities researched, 34 percent (30 out of 89) published a logic model on their websites. Of these charities, 12 came from Social Services, 11 were Children's and Family Services, and seven were Homeless Services.

Case Studies: The tenth type of information students were asked to search for and review was case studies. Case studies are defined as a detailed account of a program over a period of time that illustrates the causal link between the program's activities and a set of intended outcomes. Publishing supplementary documents further strengthens the rationale for developing a set of programs and services to address a particular issue or need. Other information intermediaries do not require charities to publish case studies to express these

relationships, but they do require that charities show the relationship between what they do and what results they achieve.

Of the 89 charities that were researched, three percent (three out of 89) published case studies on their websites. All three charities were from the Children's and Family Services sample.

Phone Survey Analysis for Element One

Element One seeks to identify how charities report on the connection between their activities and outputs and outcomes. It measures this connection with four binary tests, which are followed by a "compliance footprint" and an evaluation of "researcher confidence level." The four binary tests give a rater the option of "Yes" or "No", with no alternative option for degrees of satisfaction of test requirements.

Evidence from phone interviews with a sizable sample of charities (35 total) from each cause area confirms that the binary nature of the four tests hampers a valid assessment. Validity, in this case, refers to the ability of a test to accurately and precisely capture the true effort at results reporting by the charities it rates. The evidence compiled from the Capstone Group's phone survey suggests that, in relation to Element One, the current iteration of the CNRT may have a low degree of validity and could be strengthened through various modifications (which will be discussed in detail later).

The strongest evidence to support this conclusion comes from Supplemental Interview Question Four, which asks respondents if there are any current measurement practices that are not made public. The data generated reveals that organizations are doing more to measure and report results than is revealed on the website, and further, that the efforts are extremely

heterogeneous. They ranged from simple anecdotal interviews with clients, to pre-and post-tests, to long-term case studies. Measurement and evaluation practices included those that are less methodologically sound (staff goal-setting) to highly sophisticated and technical evaluations, requiring professional staff and advanced technology.

Overall, the responses to the phone survey suggest that charities are, in fact, measuring and evaluating programs, but that there is little standardization among them. Though discussed in more detail later, a supplemental document, designed for raters, defining the variety of appropriate ways charities might measure results will increase the validity of the CNRT by decreasing the likelihood of that raters will be confused about or overlook efforts that should be noted.

Supplemental Interview Question One, which asks about the organizational limitations to publishing results reporting measures, seeks to identify reasons why charities may utilize results reporting measures, but not report them publicly. While many organizations listed professional, technical, or resource limitations, privacy and confidentiality concerns were the most frequently offered explanations. In fact, 40 percent of organizations interviewed listed privacy of clients as a reason for not publishing results online; 17 percent specifically mentioned confidentiality concerns. One charity commented that it collected a “ream [sic] of data...but very little of it goes beyond our board.” This is not surprising, given the nature of the cause areas these organizations serve and the limited demand for results data from the average donor. The cost of publishing results, especially with respect to the fear that many organizations have of diverting funds from programming, was cited as a limitation by 26 percent of organizations. CN and the charities it rates should utilize communication channels to

mitigate these fears, as there are certainly ways that organizations can report results measurement practices without violating privacy and diverting program funds.

Interestingly, several organizations (14 percent) mentioned that they did not see the benefits of publishing results measures because it was not necessary for any reason and they believed that people did not care about these measures anyway. Expressing the opposite opinion, an even smaller number of organizations, six percent, was enthusiastic about CN 3.0's concept. Though these are minority opinions, they are indicative of the balance CN must achieve in CN 3.0: how to educate and persuade resistant charities while further encouraging those already reporting the data to do so more effectively.

Further analysis of Question One, in combination with data from Supplemental Interview Question Two, reveals that most charities are not at the extremes: the largest group of charities, 37 percent of responding organizations, made statements indicating that they disagree with CN's notion of Element One, but that if privacy/confidentiality concerns, technological issues, and organizational capacity were addressed, they could see the value of publicly reporting results measures. Thirty-one percent of organizations made statements indicating that they were "on-board" with CN's notion of Element One, albeit with some heavy reservations. (See Appendix E.2 - Supplemental Question 1.) That means that 68 percent of charities, a sizable majority, would be open to reporting results measures publicly, pending further discussion with CN. The remainder of the organizations disagreed completely with CN's point of view (11 percent), had no opinion (14 percent), or did not respond to the question (six percent).

Though subsequent chapters will address specific suggestions for improving the Element One section of the CNRT, the data presented above indicates that charities need and would appreciate resources from CN to improve their ratings. The establishment of communication channels, an advisory group, and a more knowledgeable community, outlined in the Concept Note, will be important, though CN should take precautions to prevent the development of a “teach to the test” mentality that values achieving a high rating on the CNRT over fostering performance-based management in an organization. In addition, donors will also need ways to understand the Element One component of the CNRT and how they can use the information it provides to make better informed donation decisions.

Chapter Five

Element Two — Independent Evaluations and

Standards and Certification Mechanisms

The goals of Element Two are twofold: to assess charities' efforts in publishing independent evaluations and to measure charities' practices in reporting participation in standards and certification mechanisms. As mentioned in Chapter Two, other charity information intermediaries do not specifically include independent evaluations and membership in standards and certification mechanisms as rating elements when they evaluate charities. Therefore, these two rating elements from CN 3.0 are relatively unique in evaluating charity performance. The following section will summarize major findings from the CN Research Tool (CNRT) and the Background Research Tool (BRT) regarding responses to both tests in Element Two.

The Two Tests in Element Two

Test One for this Element includes two questions which ask whether charities make "published independent evaluations" and "progress reports on corrective actions undertaken following the evaluation" available publicly. These questions are designed to encourage charities to make serious efforts to identify their program outcomes by utilizing rigorous independent evaluations and to make these program evaluation results available to the public. A charity must publish a "full, unabridged evaluation," and "the author must be financially and professionally independent of the organization," according to the CN Research Guidelines (Charity Navigator, 2012, pgs. 8, 9). These questions also attempt to ensure that charities are

adapting their programs in light of the results of the independent evaluations and that they are publishing progress reports on their corrective actions.

Test Two in the Tool includes two questions which ask whether a charity holds a “membership in one or more standards and certification mechanisms” and then, “does that standards and certification mechanism explicitly cover outcome measurement and reporting?” The purpose of these questions is to encourage charities to participate in outcome-based standards and certification mechanisms with the expectation that charities would then have their programs results reviewed rigorously and be incentivized to make progress based on their performance.

For Element Two, students filled out a total of 174 surveys for the CNRT, which covered the same three cause areas covered in Element One. Thus, the percentages provided below *for* all three cause areas combined (Social Services, Children’s and Family Services, and Homeless Services) are out of a total of 174 responses. The percentages provided for Social Services and Children’s and Family Services organizations each come from 60 student-completed online surveys, and the percentages provided for Homeless Services organizations are from 54 student-completed online surveys. Quantitative information presented in the following sections is drawn from the CNRT Frequency Tables for this Element (See Appendix D) and qualitative information is drawn from student answers to open-ended questions in CNRT and student journals written during the research process.

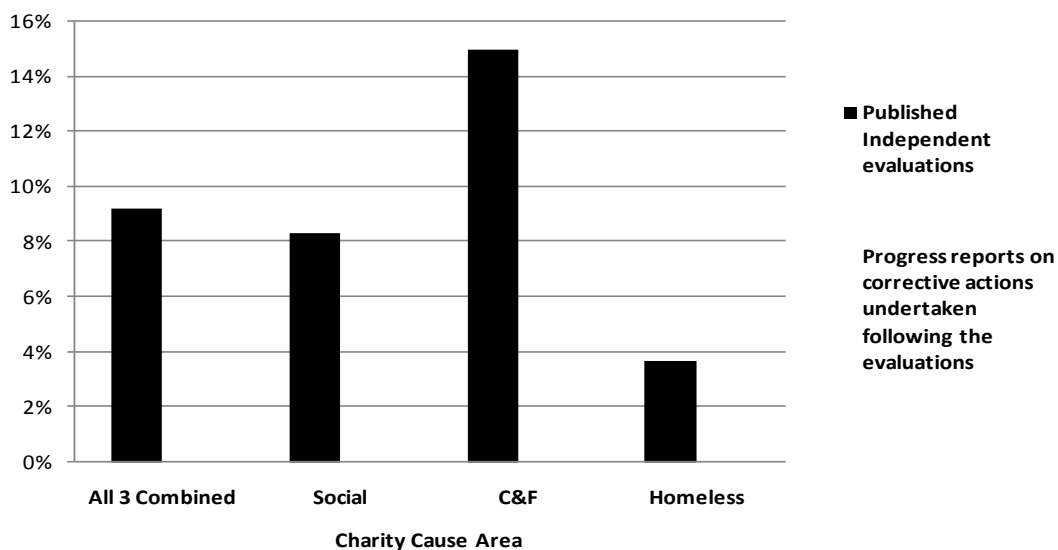
Test One: Independent Evaluations:

Quantitative: Only 16 out of 174 responses (nine percent) indicated the answer “Yes” for this question, indicating that the charity under review published independent evaluations.

Out of the 16 responses, five are from Social Services charities, nine are from Children’s and Family Services organizations, and two are from Homeless charities.

Only six out of 174 responses indicated a “Yes” to the question of whether or not they published progress reports on corrective actions undertaken following the evaluation. This accounts for 38 percent of the 16 responses that answered “Yes” to the previous question, and only three percent of the total 174 responses. Out of these six responses, none are from the Social Services organizations, five are from Children’s and Family Services organizations and one is from a Homeless Services organization. Figure 5.1 displays the percentage of “Yes” responses to both questions among the cause areas.

Figure 5.1: Percentage Breakdown of Responses to the Questions “Published Independent Evaluations” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?” in Element Two – Test One



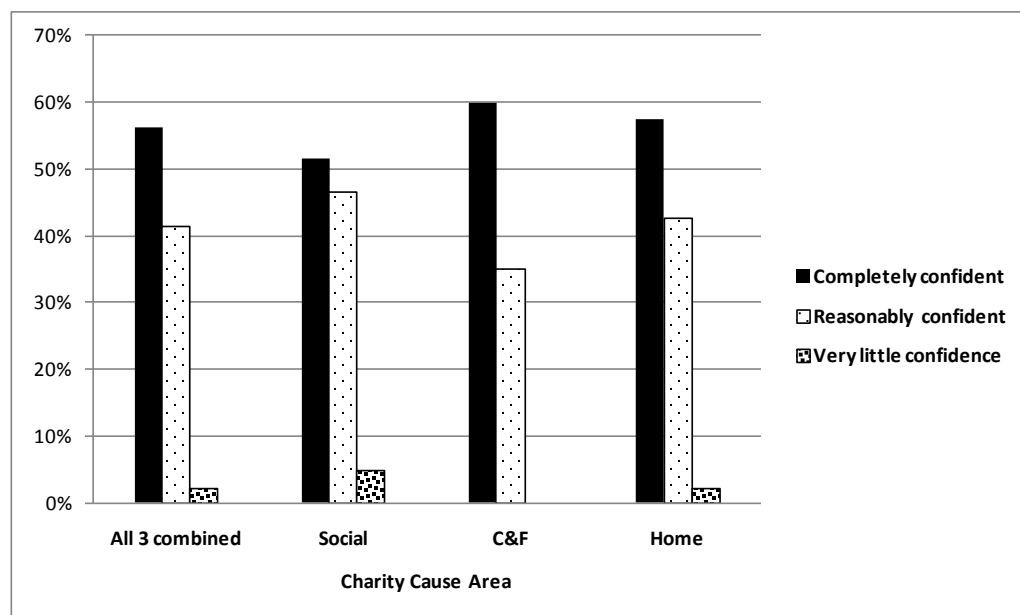
n=174

As shown in Figure 5.1 above, among the three cause areas reviewed by the SPEA Capstone Group, Children’s and Family Services organizations have the greatest number of

“Yes” responses to the question regarding publishing independent evaluations and also the greatest number of “Yes” responses to the question about undertaking corrective action following their evaluations. There were significantly fewer “Yes” responses from the Social Services and Homeless Services organizations to these two questions.

When answering these two questions, the majority (56 percent) of students indicated that they were “Completely Confident” with their answers; 41 percent of students answered “Reasonably Confident”, and only two percent felt “Not Very Confident” in responses (Figure 5.2). The average confidence level for these two questions was 2.5 out of 3 on the confidence scale.

Figure 5.2: Percentage Breakdown of Researcher Confidence Level Responses in Element Two Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?”



n=174

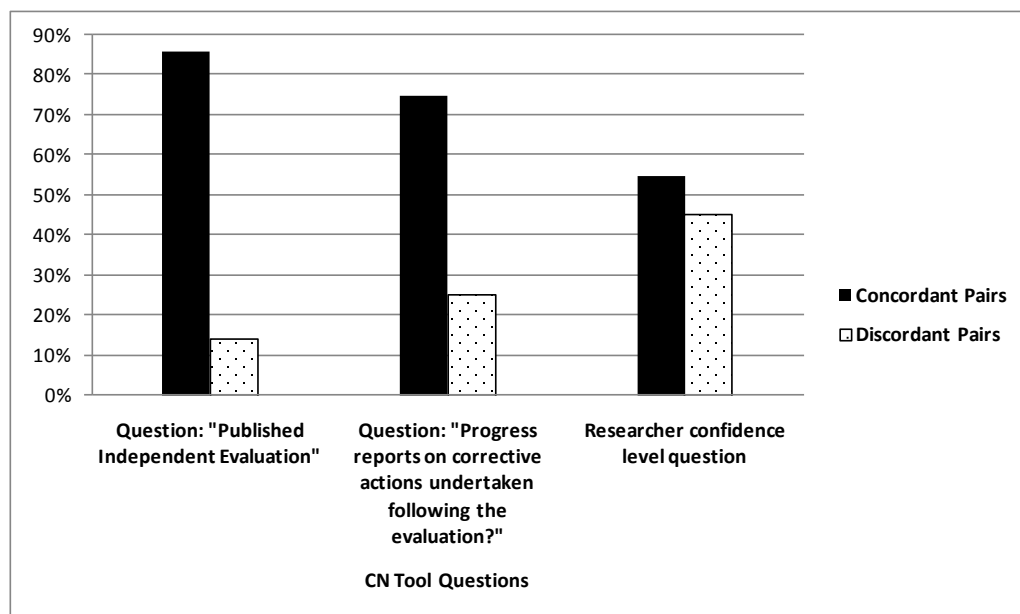
Qualitative: Students had three primary areas of concern regarding Test One: the binary nature of the questions, the level of independence of the evaluations, and confusion regarding

the confidence level questions. First, the two research questions in Test One in the CNRTI seek “Yes” or “No” answers; this prevents researchers from giving more detailed information about evaluation practices, such as the type of independent evaluation. Second, many students expressed frustration in the journals and class discussions that they were not able to determine the level of independence of the evaluations that charities undertake. For example, results from COA Youth and Family Center indicate that their programs are evaluated by their outside funders (government agencies or foundations); however, it is unclear how independent these program evaluations actually are. While students recognized the importance of independent evaluations, they also expressed the opinion that the CNRT should account for all forms of evaluation, not just completely independent ones. In addition, many students suggested that charities may need a significant amount of time to adequately transition their resource allocation to allow for independent evaluations. Finally, Test One in Element Two includes two confidence level questions separately (one for the two binary questions and one for compliance footprint question). The repetitive presence of two confidence level questions in Test One of Element Two confused many students because Element One, Element Three, and Test Two in Element Two have only one confidence level question in each of their tests.

Inter-Rater Reliability: For all three cause areas combined, the number of student pairs who gave identical answers in Test One is greater than the number of discordant pairs. For the two questions (“Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?”), the number of concordant pairs is relatively high (86 percent for the first question, and 75 percent for the second question). However, the percentage of concordant pairs for the researcher confidence level in Test One (55 percent) is

lower than those in the first two questions (Figure 5.3). Among discordant pairs for the confidence level question, the average difference in confidence level is 1.08. Among the three cause areas, on average there are more concordant pairs for each question in Social Services and Homeless Services organizations than in Children’s and Family Services organizations.

Figure 5.3: Percentage Breakdown of Inter-Rater Reliability for Element Two – Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?”



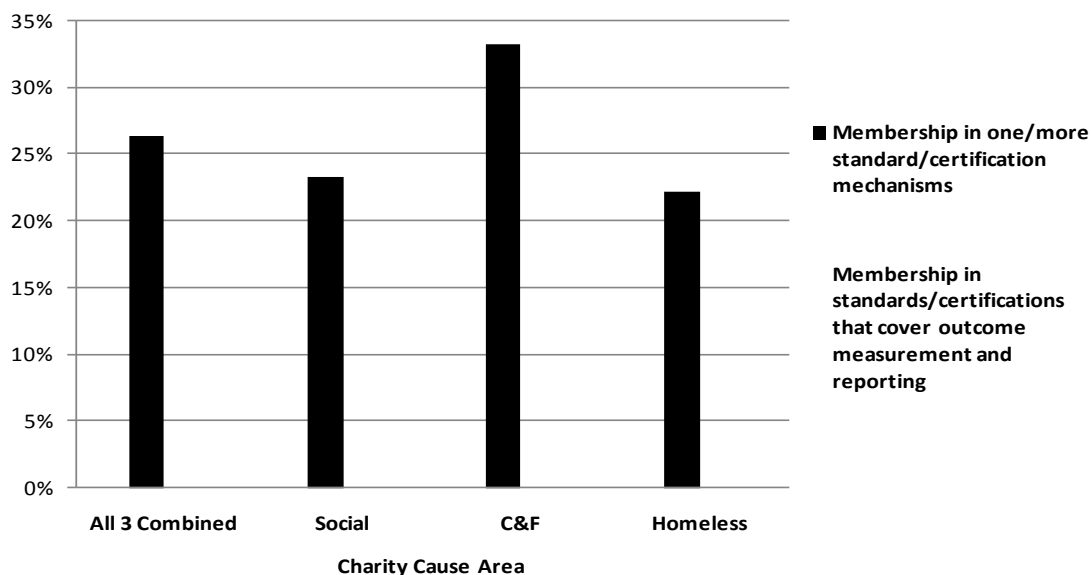
n=84

Test Two: Standards and Certification Mechanisms

Quantitative: Forty-six out of 174 responses (26 percent) were “Yes” for this question, indicating that the charity under review reported that it held a membership in at least one standards or certification mechanism. Out of the 46 responses, 14 are from Social Services charities, 20 are from Children’s and Family Services organizations, and 12 are from Homeless Services organizations.

Twenty-two out of 174 responses were “Yes” for this question, indicating that the charity under review participated in a mechanism that covers outcome measurement and reporting. This represents 48 percent of the 46 responses that answered “Yes” to the previous question, and only 13 percent of the total 174 responses. Out of these 22 “Yes” responses, two are from Social Services organizations, 14 are from Children’s and Family Services organizations and six are from Homeless Services organizations. Figure 5.4 displays the percentage of “Yes” responses to both questions among the cause areas.

Figure 5.4: Percentage Breakdown of Responses to the Questions “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?” in Test Two – Element Two



n = 174

As shown in Figure 5.4 above, among the three cause areas, Children’s and Family Services organizations have the greatest number of “Yes” responses to the question about membership in one or more standards and certification mechanisms and also the greatest number of “Yes” responses to the question about whether the mechanisms cover outcomes. Interestingly, Social Services charities indicate more mechanism memberships than Homeless

Services charities, but they have fewer outcome-based memberships than the Homeless organizations. The full list of standards and certification mechanisms that cover outcomes are analyzed under the Background Research Tool section.

There is no confidence level question for these two questions that make up Test Two, but the corresponding compliance footprint question (which is analyzed as part of Element Four in Chapter Seven) does have a related confidence level question.

Qualitative: Test Two is designed to capture charities' participation in standards and certification mechanisms. However, there are two qualitative concerns associated with this test: the challenge for the rater in adequately completing the test and the lack of a clear definition of "standards and certification mechanisms." First, many students expressed in their research journals that they felt this test was too time-consuming for the rater to complete, as its current structure requires both extensive research on a charity's website, as well as time-intensive research on the websites of the relevant standards or certification organizations to determine if they are outcome-based.

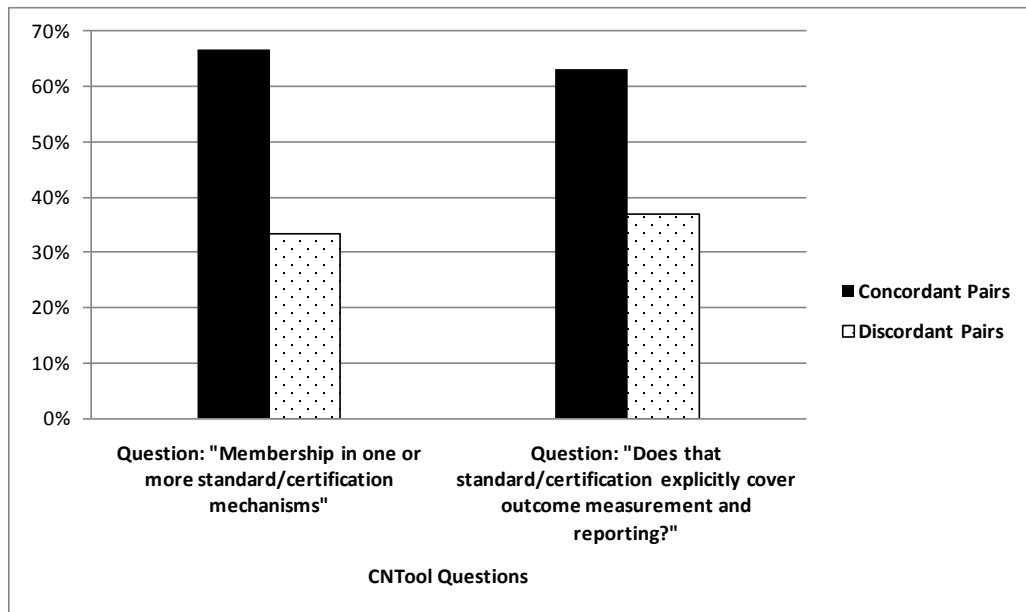
Second, this test lacks a clear definition of standards and certification mechanisms. It is unclear from the CN Research Guidelines whether "standards and certification mechanisms" cover nonprofit organizations (e.g. charity information intermediaries), public organizations (e.g. a state Department of Health), for-profit accrediting agencies, various industry-specific associations, or all of them. Students had difficulty determining what sorts of accreditation, reporting standards, or certification mechanisms were appropriate for this test.

As previously described, Test Two has one confidence level question for the compliance footprint question, but it does not have an overall confidence level question for the entire test,

like the tests in Element One and Element Three. This inconsistency was confusing for some students.

Inter-Rater Reliability: For all three cause areas combined, the percentage of students who have identical answers for the questions in Test Two is significantly greater than the number of discordant pairs (67 percent for the question: “Membership in one or more standard and certification mechanisms,” and 63 percent for the question “Does that standard and certification explicitly cover outcome measurement and reporting?”) (See Figure 5.5.) Among the three cause areas, Social Services organizations, on average, have the greatest number of concordant pairs for both questions (73 percent and 77 percent respectively), followed by Children’s and Family Services charities (67 percent and 60 percent respectively), and then by Homeless Services organizations (58 percent and 50 percent respectively). Test Two has lower inter-rater reliability than Test One, as Test Two has more discordant pairs.

Figure 5.5: Percentage Breakdown of Inter-Rater Reliability for Element Two – Test Two: “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?”



n=84

Observations from the Background Research Tool

With regard to Element Two, the Background Research Tool used two groups of questions. The first group was developed to gather information related to charities’ evaluation policies; the second group of questions explores charities’ standards and certification mechanism practices. (See Appendices D.3 and D.4 for details on the findings from the Background Research Tool in this area.)

Program Evaluation: The first group of questions asks the following:

- Is there a clear policy outlined on the website or in annual reports regarding program evaluation or performance management?
- Was the evaluation independent and who was the evaluator?
- Was the evaluation internal?
- Was a results report published, and what evaluation quality assessment tool was employed in the program evaluation?

- Was detailed information posted about how the evaluation was conducted?
- Were results posted?
- What use was made of this information?

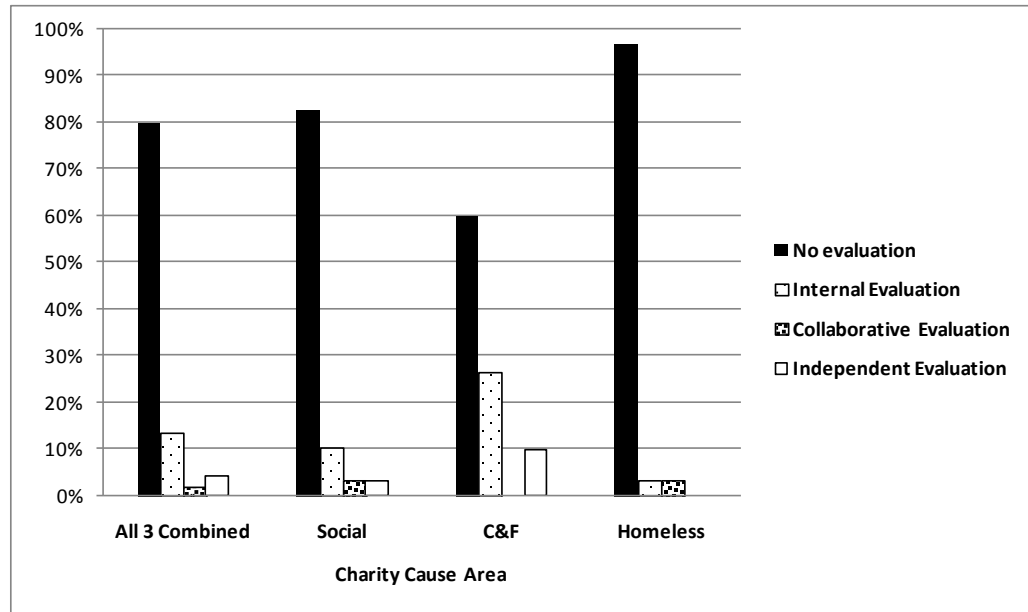
These questions were intended to help the rater determine whether organizations have evaluations of any sort, including internal evaluations, collaborative evaluations in which the evaluations are conducted by both external evaluators and charity staff, and independent evaluations. The SPEA Capstone Group recognizes the value of independent evaluations and the rationale of encouraging charities to make serious efforts to regularly conduct and publish independent evaluation results. However, considering charities' current practices in program evaluation, it may be worthwhile for donors and clients to know also about charities' efforts in other sorts of program evaluation besides explicitly independent ones. According to a 2010 study by Innovative Network, Inc., about evaluation practices and capacity in the nonprofit sector, although only 27 percent of 867 organizations worked with an external evaluator in 2009, 85 percent of 899 organizations (including the same charities as in the first study and an additional 32 organizations) had evaluated some part of their work in the past year (Reed and Morariu, 2010, p. 2-3). The CNRT asks if charities published independent evaluations and "progress reports on corrective actions undertaken following the evaluation." This group of questions aims to supplement the CN tests in order to explore more broadly charities' overall evaluation policies.

From the results of the first group of questions, students found that 18 of the 89 charities (21 percent) have evaluated their programs to some degree, publishing the results on their website. Four of the 18 charities (22 percent) conducted independent evaluations; all four of these charities published results, but only one of them had information on its website

indicating that it incorporates evaluation results into programs. Overall, only four percent of the 89 charities published their results. (A list of the reviewed charities that published independent evaluations may be found in Appendix D.) The results also show that 12 out of 18 charities (67 percent) with evaluations published internal ones; this accounts for about 14 percent of the total 89 charities. Only two of the 18 charities showed results from collaborative evaluations.

When viewed with respect to cause area, there is a slightly greater tendency for Children's and Family Services organizations to participate in and report on evaluation practices than for Homeless and Social Services organizations. Again, the most common type of evaluation was internal, with 27 percent of the 30 Children's and Family Services charities showing evidence of this type of activity. Independent evaluations were performed by about ten percent of Children's and Family Services organizations. Only about ten percent of the 29 Social Services organizations and three percent of the 30 Homeless Services organizations performed internal evaluations, and even fewer performed collaborative evaluations (Figure 5.6).

Figure 5.6: Percentage Breakdown of Evaluation Policy by Charity Cause Areas



n=89

It should be noted that the quality and quantity of information about charity participation in the evaluation process varies greatly. Some charities are explicit about their assessment methodology, partnerships with evaluators, and results. For example, Juma Ventures, a California-based nonprofit that helps impoverished youth attend college, hired BTW Consultants to perform an independent evaluation; it reported using a face-to-face interview method, published the results of the evaluation, and indicated that the results of the assessment are incorporated into its programs. Other charities merely indicated that they perform evaluations without giving much more detail, while the majority of reviewed organizations do not perform any type of assessment. Thus, the evaluation landscape is diverse and inconsistent across the field.

Standards and Certification Mechanisms: The second group of questions asks the following:

- Is there information on the website indicating that the charity participates in any of the reporting standards and/or certification mechanisms?
- Does it post logos and links to these reporting standards and/or certification mechanisms on its homepage?
- Does it have a separate section about this information on its website?
- If yes, does it provide any brief introduction about the charity's performance in these reporting standards and/or certification mechanisms?
- Did the charity win any awards or was it ranked by any agency?

The CNRT did not examine the level of transparency in the charities' participation in standards and certification mechanisms. In the initial research, students discovered that there were differences among websites regarding the way they provided information online about charities' participation in one or more mechanisms. Therefore, in the BRT, students also examined transparency in standards and certification mechanisms to see whether charities provided detailed information about their performance in relation to the guidelines of the standards and certification mechanisms in which they participate. In addition, students also listed relevant prizes or awards that charities have received since that information may also be useful for donors.

Students classified charities based on the number of standards and certification mechanisms in which charities participated, as well as ranked the standards and certification mechanisms based on the number of CN charities that participated in them in order to observe which mechanisms are the most popular among charities. The data show that the majority (78 percent) of charities under review did not participate in any reporting standards or certification

programs. Of the approximately 25 percent that did, 14 charities (16 percent of the total 89 charities) were involved with only one mechanism. Approximately eight percent of the 89 charities were involved in two or three and none participated in three or more. In total, 21 charities (24 percent) participated in 16 different standards and certification mechanisms. Among three cause areas, the BRT data show that out of 21 charities participating in standards and certification mechanisms, five are from Social Services charities, nine are from Children's and Family Services organizations, and seven are from Homeless charities.

Among these 16 standards and certification mechanisms, by far the most common was the Better Business Bureau Standards for Charity Accountability, with nine charities from all three cause areas utilizing it. The Evangelical Council for Financial Accountability (ECFA) was the second most popular, with four Homeless charities participating in it. The Independent Charities of America and National Association for the Education of Young Children (NAEYC) had three charities participating in each. (See Appendix D.)

According to the CN Research Guidelines about standards and certification mechanisms that cover outcomes, only five of these 16 mechanisms (31 percent) in which the 21 charities participated cover outcome measurement and reporting (Table 5.1). It is notable that the mechanisms mentioned above that have the highest participation rate are only process-based. They do not have any outcome-based enforcement or compliance mechanisms associated with their certification regimes.

As shown in Table 5.1, only five of the 21 charities (24 percent) participated in the five standards and certification mechanisms that cover outcome measurement and reporting. This accounts for six percent out of the total 89 charities under review. Four Children's and Family

Services charities have their programs accredited by four different accrediting agencies, and two of them (Family Centers and COA Youth and Family Center) simultaneously participate in outcome-oriented mechanisms. In another cause area, only one Social Services charity (Jewish Vocational Services) participates in such a mechanism (Commission on Accreditation of Rehabilitation Facilities - CARF). None of Homeless Services charities participated in any standards and certification mechanism that covers outcome reporting.

Table 5.1: List of Standard and certification Mechanisms that Cover Outcome Measurement and Reporting

Standard and certification Mechanisms that Cover Outcome Measurement	No. of Charities Participating	Cause Area
National Association for the Education of Young Children (NAEYC)	3	C&F
Commission on Accreditation of Rehabilitation Facilities (CARF)	1	Social Services
Community health accreditation program (CHAP)	1	C&F
Council on Accreditation (COA)	1	C&F
Wisconsin's Department of Children and Families new Youngstar rating system	1	C&F

Among the five standards and certification mechanisms that cover outcome measurement and reporting, four are voluntary accreditation programs provided by independent nonprofit organizations: CARF in the area of health and human services, CHAP in the area of healthcare, COA in the area of international child- and family-service and behavioral healthcare, and NAEYC in the area of early childhood education. NAEYC is also the most popular with three Children's and Family Services charities participating in its accreditation programs. CARF, CHAP, and COA each have one charity participating in its accreditation

programs. The last mechanism is the new Youngstar rating system designed by a public agency (Wisconsin Department of Children and Families). This system rates early childhood education and child care organizations operating in Wisconsin, either voluntarily or as a requirement of receiving state funding. One Children's and Family Services charity (COA Youth and Family Center) participated in the new Youngstar rating system. (Detailed information about how the five mechanisms cover outcome measurement and reporting may be found in Appendix D.)

The BRT also showed that only four of the 21 charities (19 percent) that participated in one or more standards and certification mechanisms have separate sections on their websites to indicate the mechanisms in which they participated and provide detailed information about how their performance was rated. The majority of other charities only posted logos and links to the standards and certification mechanisms in which they participated with no further information.

Nearly 30 percent of the 89 charities published information on their websites about prizes or awards they received from governmental agencies, media, and businesses. While there is little indication about whether these prizes or awards are outcome-based, this information may also be useful for donors.

Conclusion: Among the eight charities that either published independent evaluations or participated in standards and certification mechanisms that cover outcome measurement and reporting, six are Children's and Family Services organizations and two are Social Services ones; there are no charities from Homeless Services. (See Table 5.2) Only one charity (COA Youth and Family Center) both published an independent evaluation and participated in one or more outcome-based standards and certification mechanisms. Three charities published

independent evaluations, but did not participate in any outcome-based standards and certification mechanisms (one Social Services charity and two Children’s and Family Services charities). Four charities participated in one or two outcome-based standards and certification mechanisms, but did not publish an independent evaluation. However, COA Youth and Family Centers, a Children’s and Family Services charity, participated in two outcome-based mechanisms and had an internal evaluation. Two other Children’s and Family Services charities which participated in outcome-based mechanisms conducted evaluations that included outcome measurement: Horizons for Homeless Children indicated that it longitudinally evaluates its programs and the Jewish Social Service Agency tracks client satisfaction and publishes those results in its annual report.

Table 5.2: Comparison of Charities’ Performances in Independent Evaluation and Participation in Outcome-based Standard and certification Mechanisms

Charity Name	Cause Area	Published Independent Evaluation	Number of Outcome-based Standard/ Certification Mechanisms
COA Youth and Family Center	C&F	Yes	2
Juma Ventures	Social Services	Yes	0
Nurse Family Partnership	C&F	Yes	0
The Children’s Aid Society	C&F	Yes	0
Family Centers	C&F	No	2
Horizon's for Homeless Children	C&F	No	1
Jewish Social Service Agency	C&F	No	1
Jewish Vocational Services	Social Services	No	1

Phone Survey Analysis for Element Two

SPEA Capstone Group students conducted phone interviews with representatives of 35 charities in order to assess the external validity of the Charity Navigator 3.0 Research Tool. Through the phone interviews, students were able to determine the extent to which charities had independent evaluations and standards and certification mechanisms that were not captured by the CNRT.

Phone interviews with 11 out of 35 charities indicated they conducted some sort of evaluation that was not presented on its website. In addition, only one charity mentioned specific plans to present this information on its website in the future. The CNRT does not capture evaluations not published on the website or evaluations that are not entirely independent, though the phone interviews indicate that this is a critical aspect of the work that charities are doing. No charities mentioned reporting standards or certification mechanisms in the answers to any of the questions in the phone survey. This may reflect a general level of disinterest in this element on the part of charities.

In addition, students asked charities to discuss limitations to publishing the information relevant to Element Two on websites. Eight out of 35 charities mentioned cost, which applies specifically to Element Two, as any kind of evaluation can be expensive. In addition, eight of the 35 charities interviewed were concerned about the utility of the information to donors, as evaluation results can be complex and not all charities believe that donors demand quantitative data. Some charities conduct evaluations in line with procedures determined by a grant, and these charities felt that this format was difficult to present to the public and not worth the resources necessary to do so.

Students also asked charities to speak about their perception of Element Two's value. Though conducting independent evaluations may be an important criterion for Charity Navigator, it appears that most charities do not see them as a high priority in light of resource constraints. All comments on Element Two mentioned the limitations of independent evaluations. Five of the 35 charities interviewed indicated that independent evaluations are too time- and resource-intensive for the benefits that they provide. In addition, charities mentioned that independent evaluations might not be appropriate for every organization. Two charities mentioned that they hire their own research staff specifically to conduct internal evaluations, rather than utilizing an independent entity.

Overall, given the results of the phone survey, it appears that the CNRT does not capture the extent of evaluations that the charities selected by CN conduct and publish, seemingly because many of the evaluations are not explicitly independent. Though there are resource constraints, it does seem that most charities find value in evaluations. In addition, it is clear that what charities post on their websites is not necessarily indicative of the work they are doing in this area. However, it does not appear that many of the charities reviewed will be posting evaluations (independent or otherwise) in the near future. More communication with charities is needed to make them fully understand the purpose of CN 3.0 in this area and encourage them to publish evaluation results, especially those that are not completely independent.

Chapter Six

Element Three — Constituent Voice

Though the for-profit sector has long been concerned with client satisfaction, the concept of constituent voice presents a new and different way of thinking about it. Charities have multiple constituent groups and stakeholders to whom they are accountable. Donors can have their voices heard because of donor demands and strings attached to funding, which make the charities pay attention to funders' requests. Staff and volunteers' voices can be taken into consideration as charities work to recruit and retain them for boards and operations. But one of the constituent groups that may often be overlooked is primary constituents, or direct recipients of service. Element Three of CN 3.0 encourages charities not only to collect client feedback, but also to report it to the public and the primary constituents, so that charities can improve programs and service delivery to better meet the needs of those they are trying to serve.

The Six Tests for Element Three

Element Three has six tests. The first asks "Does the charity publish rigorously collected feedback data from its primary constituents?" The intent behind this question is to ensure that charities are engaging in substantive feedback data collection. For such collection to occur, charities need to collect data for more than marketing purposes, leading to more than "a collection of selected stories about people helped," as mentioned in the instructions. Also, consistency is a major theme behind this question, as charities are asked to use the same questions and collect the data the same way repeatedly. This question attempts to respond to

the conclusion from *The 21st Century Potential of Constituency Voice (21st Century)* that “feedback is underdeveloped and therefore of inconsistent value” (Bonbright, Campbell, & Nguyen, 2009, p. 23). The CN instructions clarify that the focus of this Element should be the constituency group of clients or direct service recipients of the charity. Feedback from other possible constituents does not count.

Test Two in the Charity Navigator Research Tool for constituent voice asks, “Is that published feedback data presented in a way that shows changes over time going back at least one year?” The intent behind this question is that creating consistent, measurable results over time allows a charity to learn from its history what programs were more satisfying to clients when implemented in a certain way, which helps with strategy development and program implementation. Measuring multiple iterations of program quality, as seen by clients, creates a trend-line that that charities can then analyze.

Test Three in the CNRT for constituent voice asks, “Is there a description of how much the published data can be considered to be representative of all primary constituents?” The intent of this test is for a charity to be forthcoming with its sample size, demographics, and other characteristics that would describe how representative the feedback data are. This test contributes to the accountability and transparency goals of results reporting, specifically with regard to constituent feedback. Though the instructions allow raters to infer representation from different locations on the charity’s website, the more transparent a charity is about the characteristics of its data, the higher value it is deemed to place on constituent voice.

Test Four of the CNRT asks, “Does the feedback data include questions that speak to the organization’s effectiveness?” The intent of this question is to assess if the feedback collection

method is related to the improvement purpose for feedback, according to *21st Century*, which “emphasizes the primary value of feedback to provide organization leaders with information for strengthening performance, maximizing critical management values, such as efficiency, improving organizational learning and assessing the overall effectiveness and impact of a charity’s work” (Bonbright, Campbell, & Nguyen, 2009, p. 16). The feedback data on the website need to include the questions asked for a charity to pass this test. Because effectiveness questions asked by the charity can look very different, raters may have to infer the purpose of the question asked.

Test Five of the CNRT asks, “Does the organization report back to its primary constituents what it heard from them?” The intent of this question is to assess the commitment of management to be accountable to its primary constituents when using their feedback, encouraging charities to “adopt the feedback research question for public reporting,” according to *21st Century* (Bonbright, Campbell, & Nguyen, 2009, p. 34). This test suggests that charities should not only return feedback results to the providers of the data, but also communicate that data to the overall community to create an expectation for the charity to consistently ask for feedback. It also creates charity accountability by creating the expectation in the community that programs or services are going to be improved based upon the feedback. This expectation of continual feedback and program improvement based on feedback makes primary constituents feel valued, encouraging constituents to provide more quality feedback in the future as trust develops, and allows the charity to continue to foster constituent loyalty and satisfaction in service delivery.

Test Six of the CNRT asks, “Does the organization benchmark its feedback data against other organizations?” The intent of this question is to make charities aware of the importance of creating comparative feedback across charities “which simultaneously clarifies the meaning of feedback data by establishing a norm, and creates incentives to take feedback seriously in order to be best in class,” according to *21st Century* (Bonbright, Campbell, & Nguyen, 2009, p. 34). Creating this norm involves developing questions based upon another charity’s feedback questions, and then comparing the feedback using another charity as a baseline. Doing this should provide charities with clearer ways to interpret high or low scores, and gives primary constituents the chance to see which charity has higher client satisfaction.

Students completed 174 surveys using the CNRT, incorporating information from charity websites from all three cause areas; thus, the percentages below are reported in the aggregate and for the three separate cause areas. For total percentages, it was necessary to create two sections: one section comparing the number of “Yes” responses to the entire 174 responses, and a second section comparing the number of “Yes” responses to the number of responses that completed the threshold question of Test One for the constituent voice element (31 observations out of the 174 total observations). (See Appendix E for Constituent Voice Frequency Tables.)

Test One: Does the charity publish rigorously collected feedback data from its primary constituents?

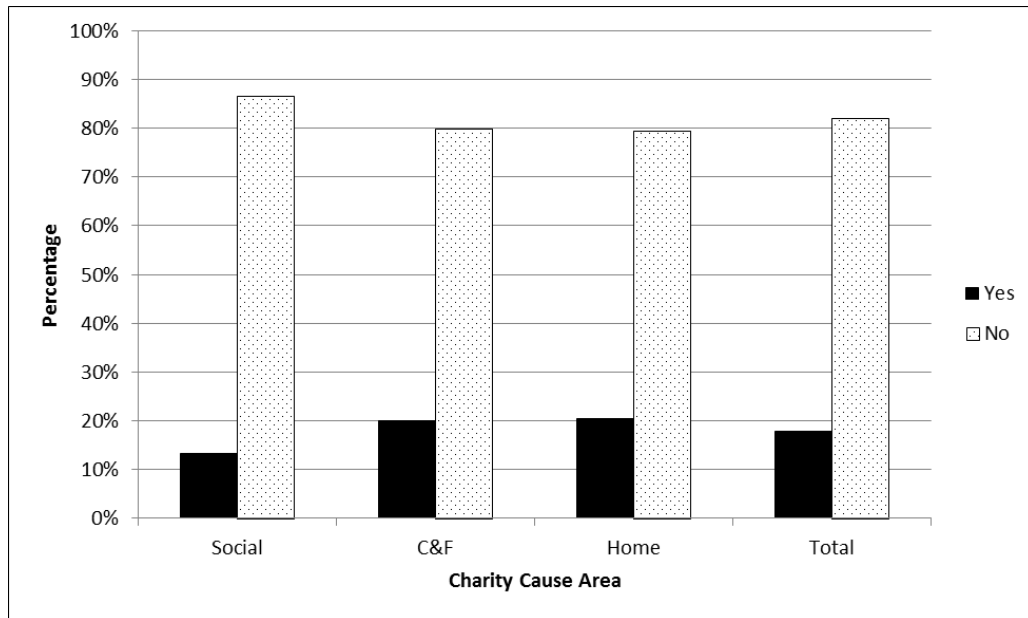
Quantitative: Only 31 student-collected responses out of 174 (18 percent) chose the answer “Yes” for Test One. Out of the 31 responses, eight are from Social Services charities, 12 are from Children’s and Family charities, and 11 are from Homeless Services charities. Figure

6.1 displays the percentages of “Yes” responses among the cause areas. The majority (52 percent) of students indicated that they were only “Reasonably Confident” by choosing the confidence level of two. As Figure 6.2 shows, the confidence level of “Completely Confident” was chosen 42 percent of the time, while “Very Little Confidence” was chosen only six percent of the time.

On Test One, 20 percent of both Homeless Services and Children’s and Family Services charities were found to do rigorous collection of feedback data. Of Social Services charities, only 13 percent had rigorous collection of feedback data. The Capstone Group had a very high level of overall confidence, even though most answers were “no.” Among students, inter-rater reliability was 73 percent, showing that even though most students were confident in choosing their answer, they did not necessarily agree on what the standard for “rigorous” was and pairs were frequently discordant (Figure 6.3).

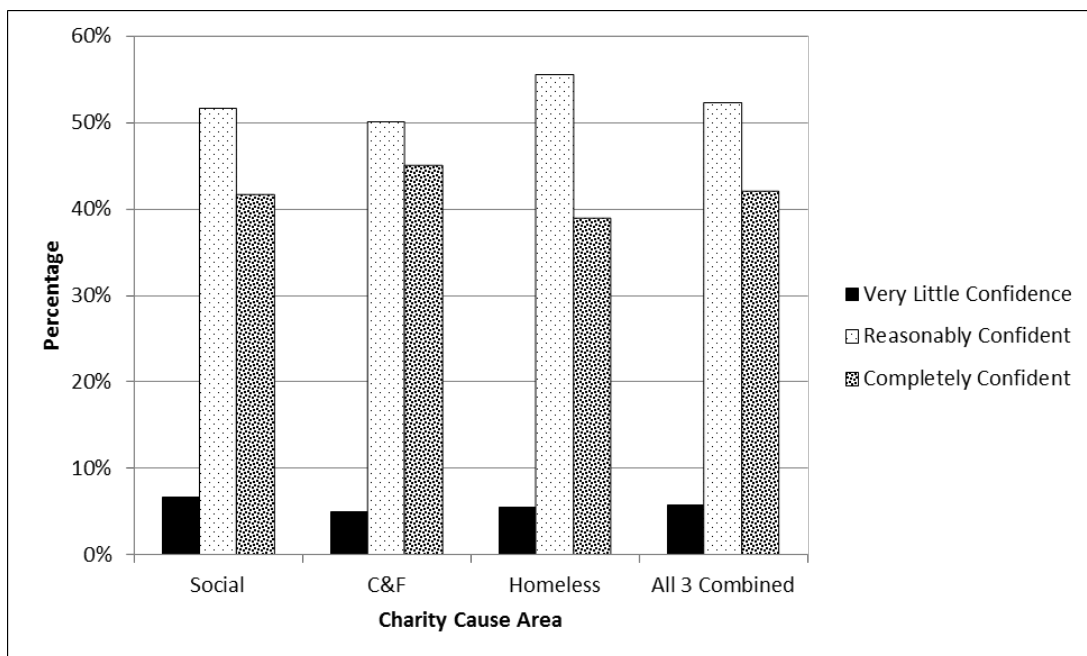
For the 31 students who responded “Yes,” three students (10 percent) chose “Very Little Confidence,” while the 28 remaining students (90 percent) equally chose “Reasonably Confident” and “Completely Confident” (Figure 6.4). For the 143 students who responded “No,” seven students (five percent) chose “Very Little Confidence,” 77 students (54 percent) chose “Reasonably Confident,” and 59 students (41 percent) chose “Completely Confident.” Since 95 percent of students were at least “Reasonably Confident” in responding with “No” to the threshold test and 90 percent were at least “Reasonably Confident” in responding with “Yes,” students, in general, were marginally less confident in concluding that the charity did in fact rigorously collect feedback on its primary constituents. Possible reasons for these results will be discussed in the following sections.

Figure 6.1: Percentage Breakdown of Element Three – Test One: Does the charity publish rigorously collected feedback data from its primary constituents? Binary Question



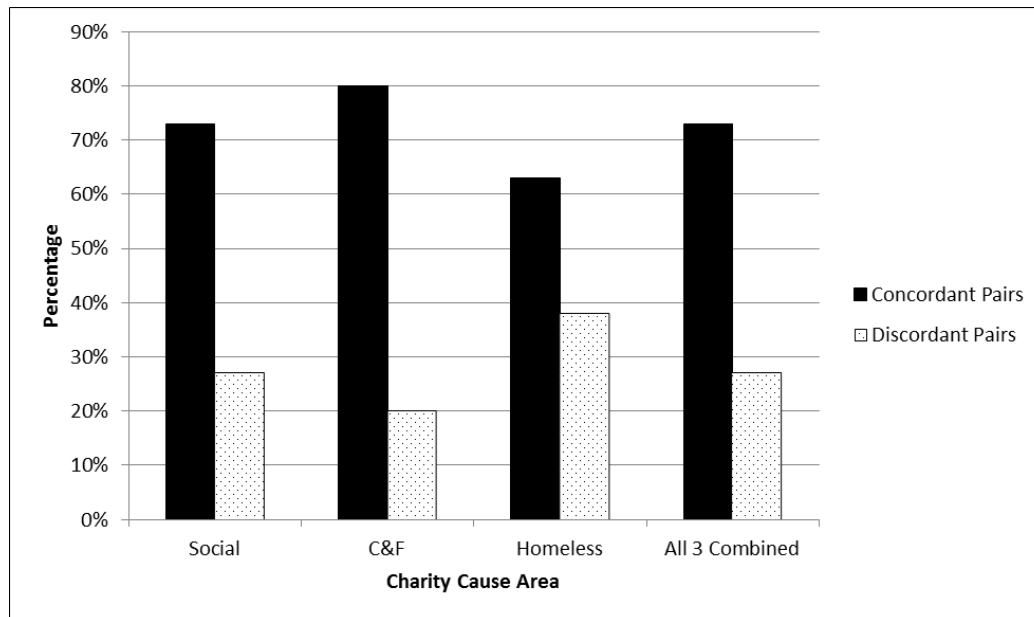
n=174

Figure 6.2: Percentage Breakdown of Element Three – Test One: Does the charity publish rigorously collected feedback data from its primary constituents? Researcher Confidence Level Responses



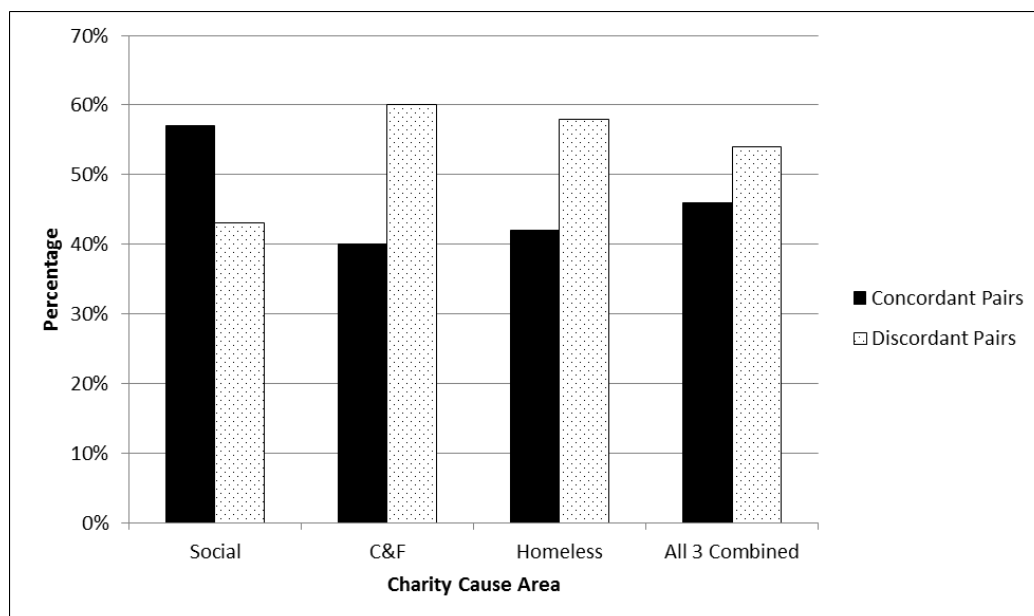
n=174

Figure 6.3: Percentage Breakdown of Inter-Rater Reliability for Element Three – Test One: Does the charity publish rigorously collected feedback data from its primary constituents? Binary Question Responses



n=84

Figure 6.4: Percentage Breakdown of Inter-Rater Reliability for Element Three – Test One: Does the charity publish rigorously collected feedback data from its primary constituents? Research Confidence Level Responses



n=84

Qualitative: Students wrote in the journals that Test One needed clarification in three major areas: subjectivity, the binary nature of the answer, and the stop-gate format of the question. First, many students felt the phrase “rigorously collected” was poorly defined, leaving the reader with too much room for interpretation. Many journals mentioned students had a difficult time discerning the degree of rigor of feedback collection mechanisms; for instance, is an interview with a primary constituent considered rigorous? Some students thought it did not, while others thought it did, if repeated with more than one interviewee. Furthermore, a binary answer did not allow students the opportunity to say how rigorous (very rigorous, somewhat rigorous, and not rigorous) the feedback collection process was. Finally, because a “No” answer did not allow students to answer the remaining constituent voice questions, some aspects of constituent voice did not get recorded in the CNRT for many charities. There are aspects of constituent voice, such as client testimonials, that a charity could acquire and use without having a rigorous system in place to collect or analyze the data.

Students were also unsure as to how far into the various links and publications from a charity they should have searched when looking for the published feedback on the website. Some feedback is listed in annual reports, while other organizations list it on the front page of the website. This made it difficult to know where to find rigorously collected feedback information and left some students searching extensively through various publications, while others did not review publications at all.

Inter-Rater Reliability: Overall, there were 84 organizations reviewed by two people each using the CNRT. Of the responses, 23 pairs of students answered Test One of the CN Tool with different responses, making them discordant, i.e. one person put “yes” while the other put

“no.” This indicates that there was 73 percent inter-rater reliability, with 27 percent of the pairs having differing answers.

Since Test One was a threshold question, only four of the pairs agreed and gave a “yes” response with the CNRT, allowing them to move on to Tests Two through Six. This would mean that of the 61 pairs with matching answers, only seven percent had both individuals putting “Yes.” Since the number of responses for Tests Two through Six was so small, showing the inter-rater reliability of these test responses would not be meaningful.

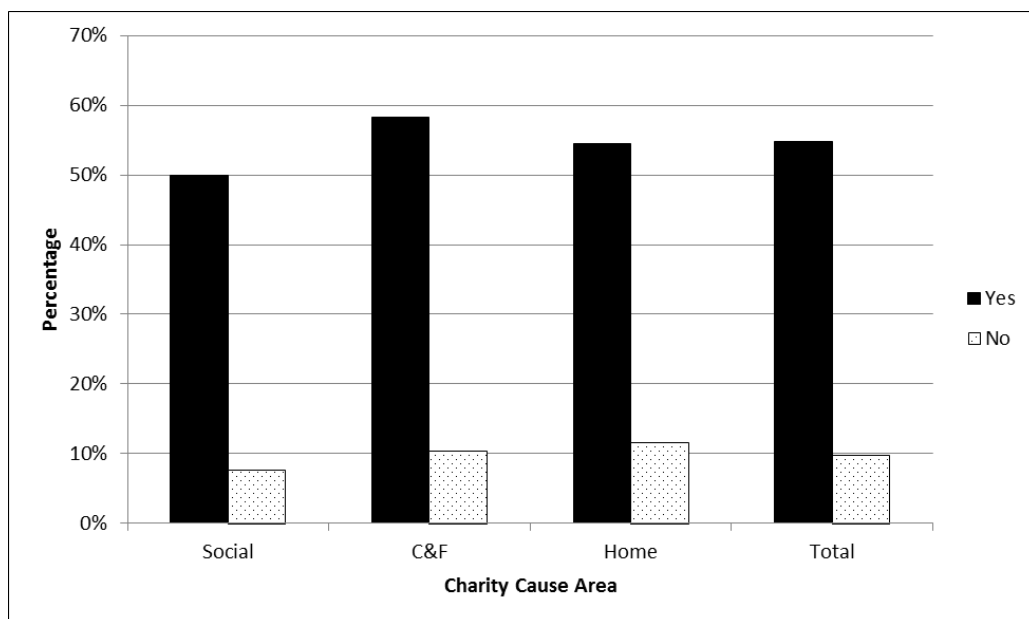
With regard to the Confidence Level for Test 1, a majority of the pairs, 54 percent, did not agree with each other while 46 percent did. Once again, due to the low number of pairs passing the threshold question, the inter-reliability results of Confidence Levels for Tests Two through Six are not meaningful.

Test Two: Is that published feedback data presented in a way that shows changes over time going back at least one year?

Quantitative: For Test Two, 17 students responded “Yes” to this question, comprising 55 percent of the 31 responses that reached this question, and comprising 10 percent of the total 174 responses collected. Of these 17 “Yes” responses, four were from Social Services charities, seven were from Children’s and Family Services charities, and six were from Homeless Services charities. Figures 6.5 and 6.6 below show the percentages of the responses in each cause area. When answering these questions, as Figure 6.7 shows, the majority (52 percent of students), stated they were “Completely Confident” in their answer. Also, 42 percent of students felt “Reasonably Confident,” and six percent felt “Very Little Confidence” in responses.

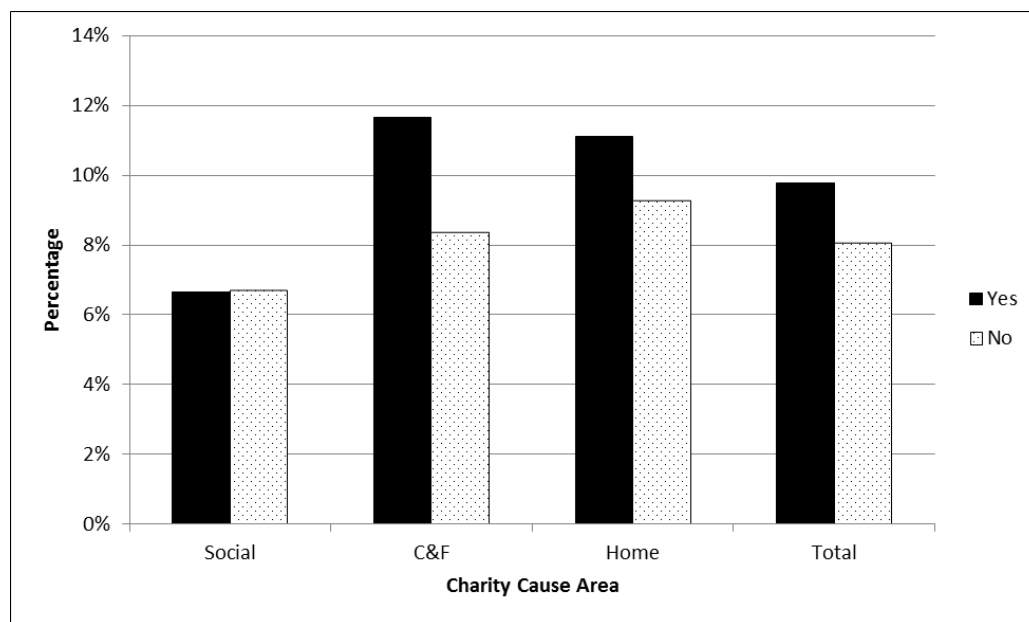
Overall, 55 percent of charities that made it to this question had available feedback data that went back over one year. Within the different types of charities, this percentage varied little. Answer confidence for this test varied greatly, with more students selecting “Reasonably Confident” than “Completely Confident” for Homeless Services charities, as opposed to other types of charities, giving the category the lowest confidence rating overall.

Figure 6.5: Percentage Breakdown of Element Three – Test Two: Is that published feedback data presented in a way that shows changes over time going back at least one year? Binary Question (“Yes” responses to Test One)



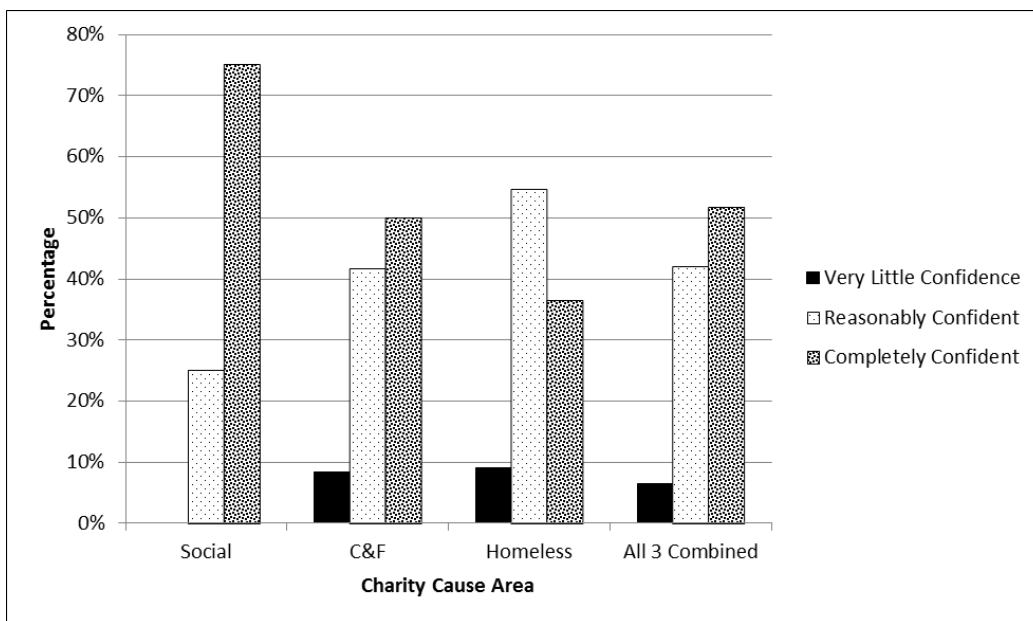
n=31

Figure 6.6: Percentage Breakdown of Element Three - Test Two: Is that published feedback data presented in a way that shows changes over time going back at least one year? Binary Question (Total sample)



n=174

Figure 6.7: Percentage Breakdown of Element Three – Test Two: Is that published feedback data presented in a way that shows changes over time going back at least one year? Researcher Confidence Level Responses



n=31

Qualitative: Test Two's primary concerns dealt with the intent of the question and difficulty charities have meeting this requirement. First, the question's wording created ambiguity. The test asks whether or not the published feedback data shows changes over time going back at least one year. However, the question's instructions do not reiterate a time period, but state that "a survey repeated only once would meet the test." With these instructions, one could infer that a test done once twenty years ago could meet this standard, while a test conducted monthly or quarterly for only nine months would not meet this standard, despite repeated testing. This led to student uncertainty, as reflected in the journals.

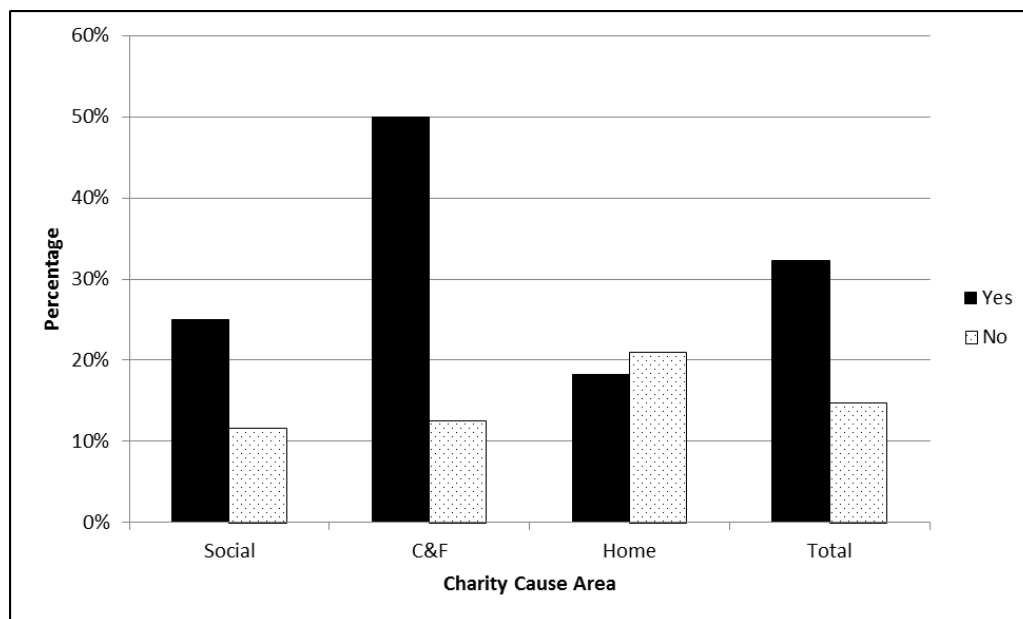
Also, certain types of charities might find it difficult to collect long-term data from clients who change regularly, such as residents of homeless shelters. This turnover between testing periods can affect results and make drawing conclusions from long-term testing hazardous.

Test Three: Is there a description of how much the published data can be considered to be representative of all primary constituents?

Quantitative: For Test Three, only ten students answered "Yes," comprising 32 percent of the 31 responses that reached this question, and six percent of the 174 total responses. Of these ten responses, two were for Social Service charities, six represented Children's and Family Services, and two were from Homeless Services. The percentages in each cause area are shown in Figures 6.8 and 6.9. As seen in Figure 6.10 below, the largest share of respondents, at 45 percent, felt "Reasonably Confident." Also, 42 percent of the students said they were "Completely Confident" and 13 percent of the students responded with "Very Little Confidence."

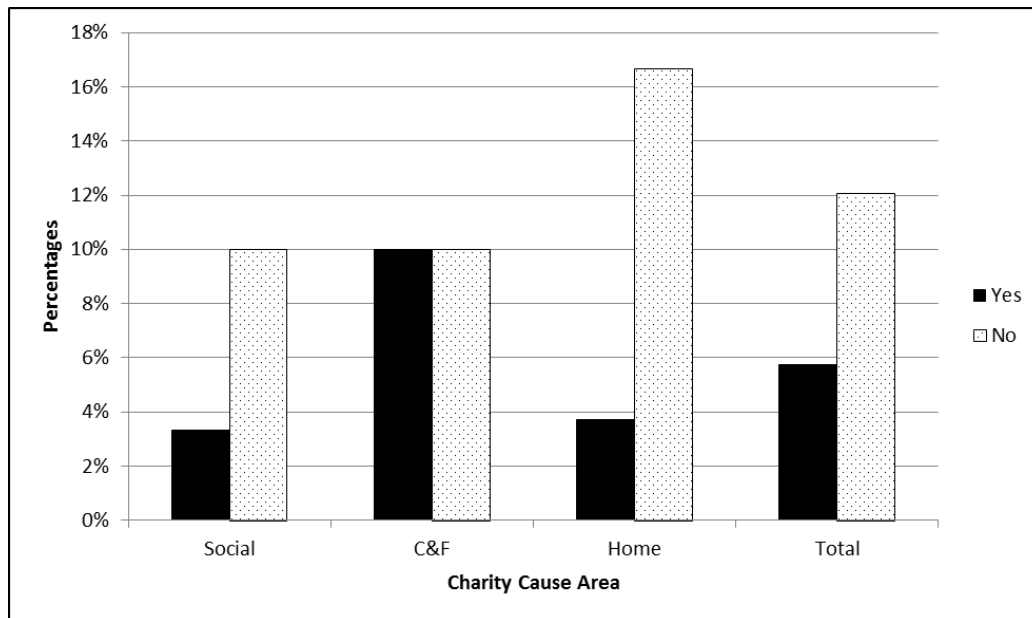
For Test Three, 32 percent of the charities assessed had information concerning how representative published results were. Charity fields varied widely, with 50 percent of Children’s and Family Services charities indicating client feedback data were representative, and only 18 percent of Homeless Services charities doing so. Confidence scores underscore this, as Homeless Services charities had the highest percentage of students (27 percent) having little confidence in their answers, 14 percentage points more than the average and 19 percentage points above the next closest category of organization. This could show confusion in student evaluations of what “representative” means and the particular difficulty Homeless Services charities have in showing representative sampling.

Figure 6.8: Percentage Breakdown of Element Three – Test Three: Is there a description of how much the published data can be considered to be representative of all primary constituents? Binary Question (“Yes” responses to Test One)



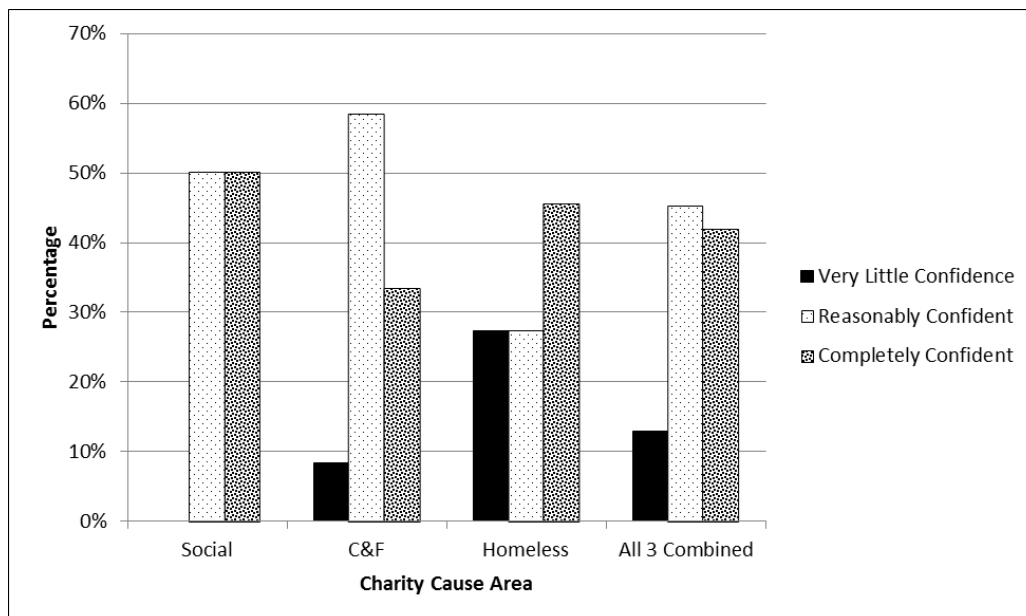
n=31

Figure 6.9: Percentage Breakdown of Element Three – Test Three: Is there a description of how much the published data can be considered to be representative of all primary constituents?
Binary Question (Total sample)



n=174

Figure 6.10: Percentage Breakdown of Element Three – Test Three: Is there a description of how much the published data can be considered to be representative of all primary constituents? Research Confidence Level Responses



n=31

Qualitative: The only qualitative concern with Test Three was the meaning of the test. CN 3.0 attempts to focus on publically reported information, not the quality of information reported. To the students, this question seemed to request a judgment on the reported information and whether or not it was a valid representation of constituents' population size, demographic representation, and other characteristics. The question would benefit by being more explicit about this distinction and a statement of "representativeness" is sufficient or more documentation is required.

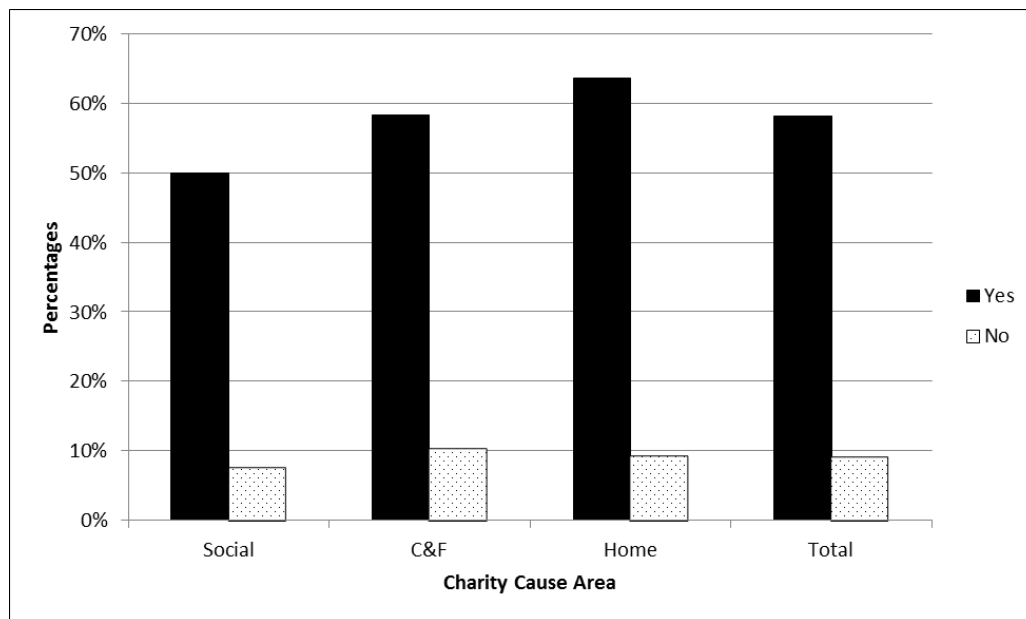
Test Four: Does the feedback data include questions that speak to the organization's effectiveness?

Quantitative: For Test Four, 18 of the students who answered this question responded with "Yes," making up 58 percent of the 31 responses that reached this question, and ten percent of the 174 total responses. Of these 18 responses, four came from Social Services charities, seven from Children's and Family Services, and seven from Homeless Services. The percentages in each cause area are shown in Figures 6.11 and 6.12. As shown in Figure 6.13 below, the largest percentage of students, at 48 percent, responded they were "Reasonably Confident." Forty-two percent said they were "Very Confident" and ten percent had "Very Little Confidence."

There was some variability among organizational categories, with Homeless Services charities having the highest percentage of "Yes" answers (64 percent) and Social Services charities having the lowest (50 percent). Again, confidence varied among the different types of charities, with Homeless Services charities producing the least confident answers. Within this category, 18 percent had "Little Confidence" in their answers, eight percentage points more

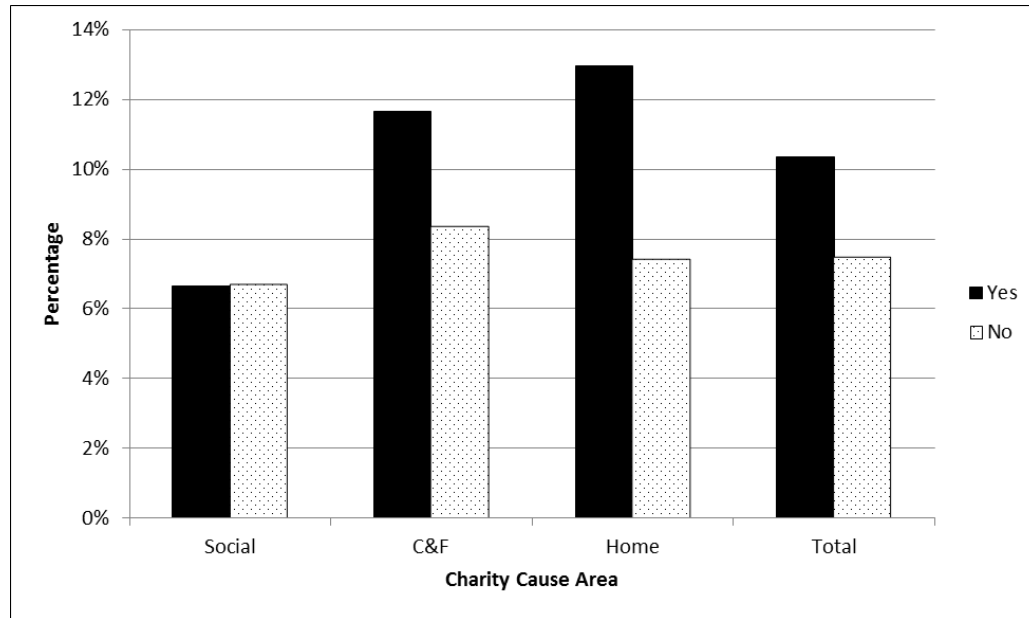
than average. Likewise, only 27 percent of responses were “Completely Confident,” which is 15 percentage points below average.

Figure 6.11: Percentage Breakdown of Element Three – Test Four: Does the feedback data include questions that speak to the organization’s effectiveness? Binary Question (“Yes” responses to Test One)



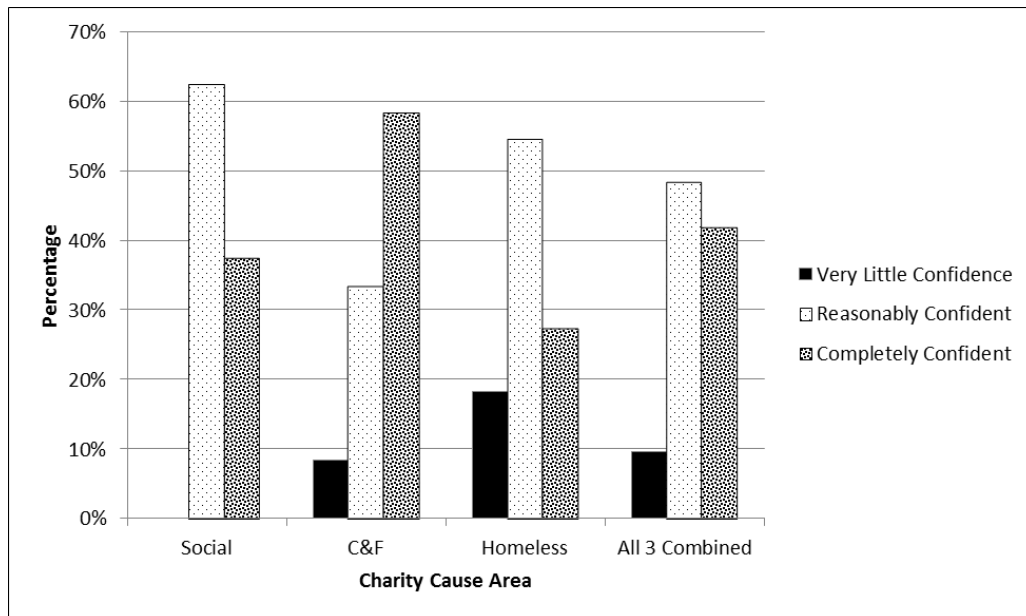
n=31

Figure 6.12: Percentage Breakdown of Element Three – Test Four: Does the feedback data include questions that speak to the organization’s effectiveness? Binary Question (Total sample)



n=174

Figure 6.13: Percentage Breakdown of Element Three – Test Four: Does the feedback data include questions that speak to the organization’s effectiveness? Researcher Confidence Level Responses



n=31

Qualitative: For Test Four, students expressed concern about the clarity of the question. Though the majority of students were confident in their answers, some did not understand what “questions” were to be rated, as the term could encompass interview questions, survey questions, social media forums, etc. It may be helpful to explain within the instructions whether “feedback data” may consist of surveys, forums, or other question-filled materials, or, a more limited term may be more appropriate.

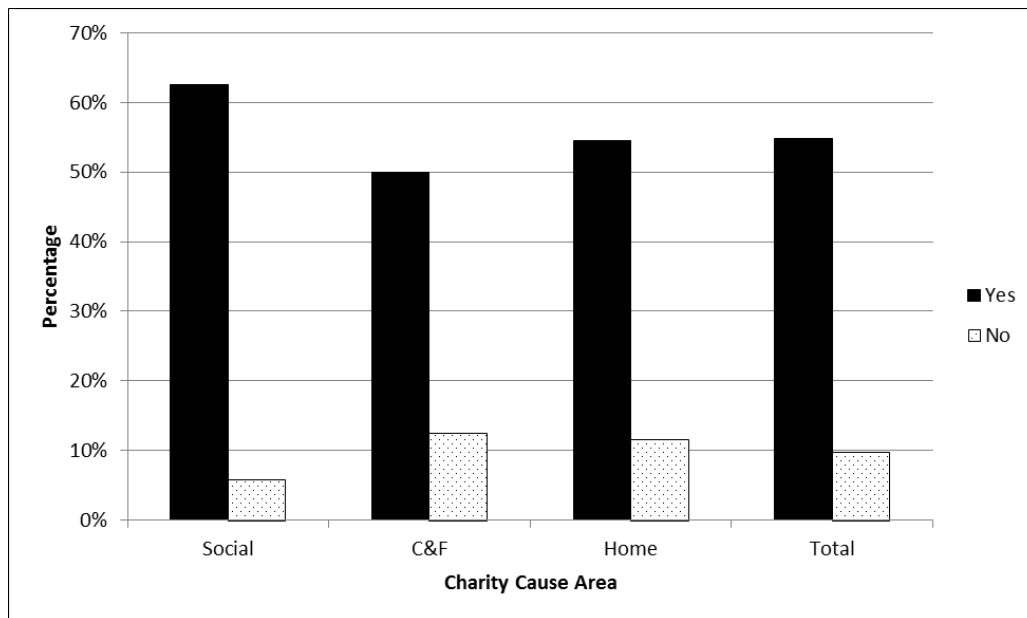
Test Five: Does the organization report back to its primary constituents what it heard from them?

Quantitative: On Test Five, 17 students were able to respond “Yes,” comprising 55 percent of the 31 responses that reached this question, and ten percent of the 174 total responses. Among these 17 responses, five were from Social Services charities, six from Children’s and Family Services, and six from Homeless Services. The percentages in each cause area are shown in Figures 6.14 and 6.15. Figure 6.16 indicates the largest share of the students (48 percent) felt “Reasonably Confident” when answering this question. Of the other students, 32 percent felt “Very Confident” and 19 percent felt “Very Little Confidence.”

There was also variability in the “Yes” answers for Test Five. Between the highest category, Social Services charities, and the lowest category, Children’s and Family Services charities, there was a 13-point spread, from 50 to 63 percent answering “Yes.” The differences in confidence between the three categories was similar to Question Four, with Homeless Services Organizations receiving the least confident evaluations, especially in the “Completely Confident” answer, which was 20 percentage points below the next lowest category group of charities. A peculiarity of Children’s and Family Services charities was that the confidence

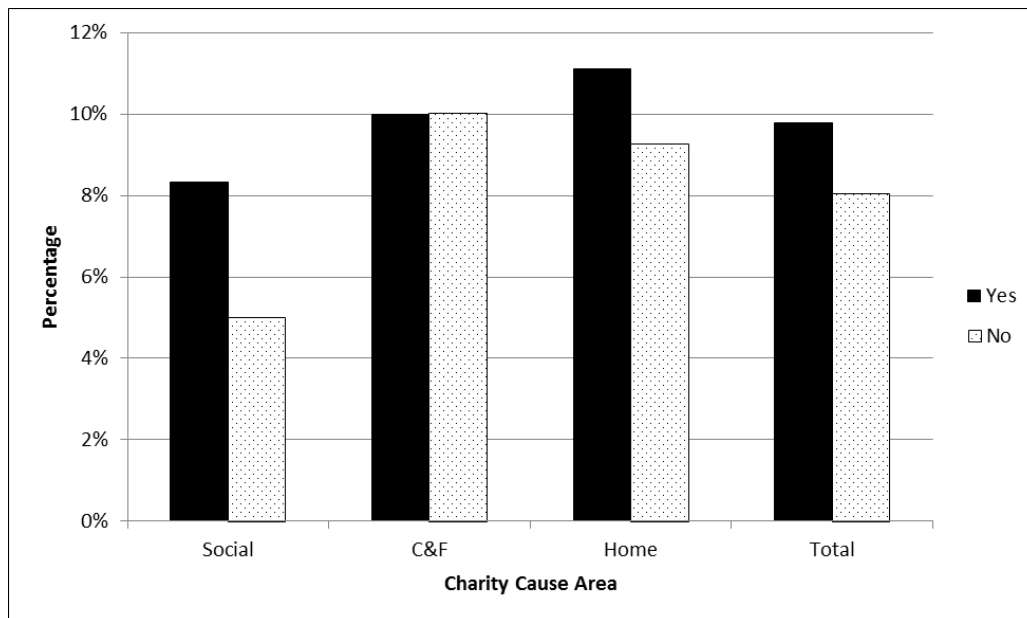
spread was almost equally divided between the three answer choices, indicating only moderate confidence in answers about this question.

Figure 6.14: Percentage Breakdown of Element Three - Test Five: Does the organization report back to its primary constituents what it heard from them? Binary Question (“Yes” responses to Test One)



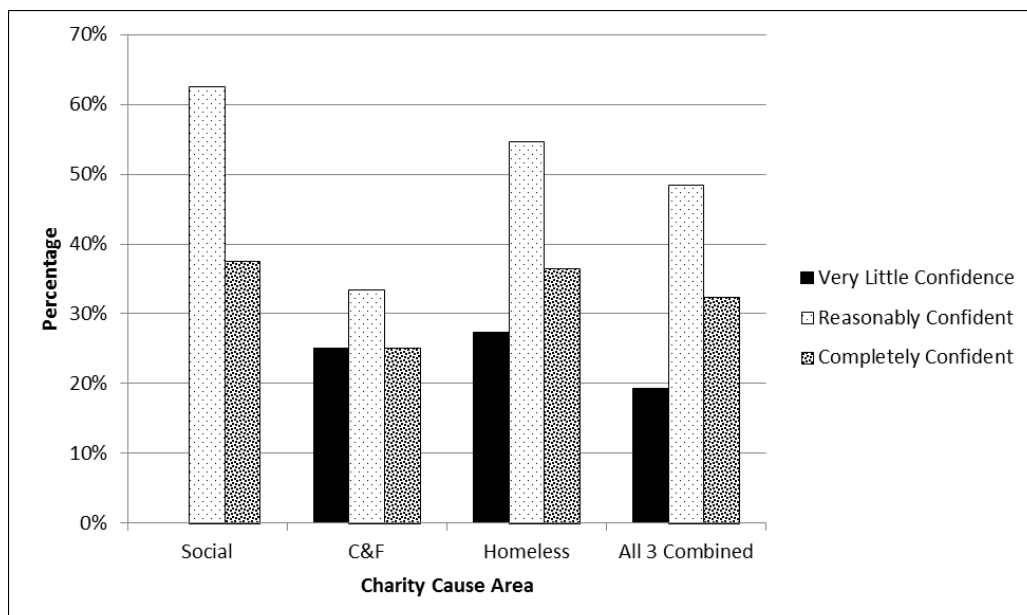
n=31

Figure 6.15: Percentage Breakdown of Element Three - Test Five: Does the organization report back to its primary constituents what it heard from them? Binary Question (Total sample)



n=174

Figure 6.16: Percentage Breakdown of Element Three - Test Five: Does the organization report back to its primary constituents what it heard from them? Researcher Confidence Level Responses



n=31

Qualitative: When completing Test Five, there seemed to be some confusion as to what was considered “reporting back” to primary constituents, as some students thought that a charity posting the information on the website could meet this test, while others did not. One needed point of clarification is whether posting the results of surveys online would be considered reporting back or whether the information needed to be shared in a specific way. Also, in Children’s and Family charities, the primary constituents need to be clarified, as it could be either the children or the parents.

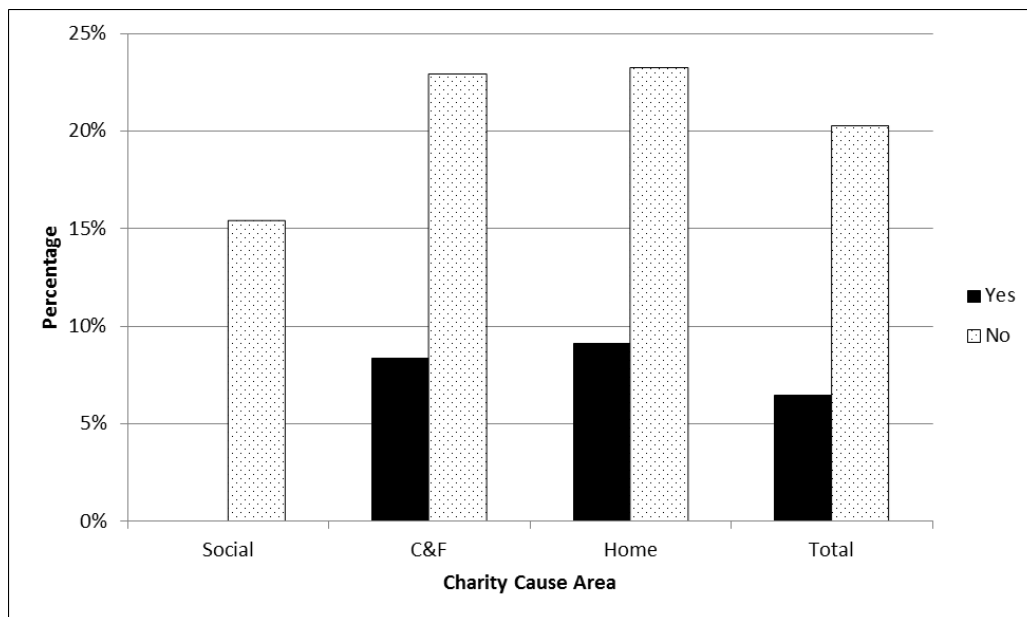
Test Six: Does the organization benchmark its feedback data against other organizations?

Quantitative: Only two students were able to respond “Yes” to the question, comprising six percent of the 31 responses that reached this question and one percent of the 174 total responses. Out of the two responses, one was from Homeless Services, while the other was from Children’s and Family Services. The percentages are shown below in Figures 6.17 and 6.18. As seen in Figure 6.19, among the 31 responses, a majority of the students (45 percent) indicated they felt “Completely Confident.” Of the remaining students, “Reasonably Confident” was chosen by 39 percent and “Very Little Confidence” was chosen by 16 percent.

Because very few organizations benchmark their own feedback data results against those of other organizations, the number of “yes” answers was very low, involving less than 10 percent of Children’s and Family Services and Homeless Services charities’ websites. Confidence in answers was relatively high, despite the lack of “yes” answers, with the greatest certainty existing in Social Services charities (88 percent), largely because none benchmarked. The other two categories still had high confidence levels, with 83 percent of Children’s and

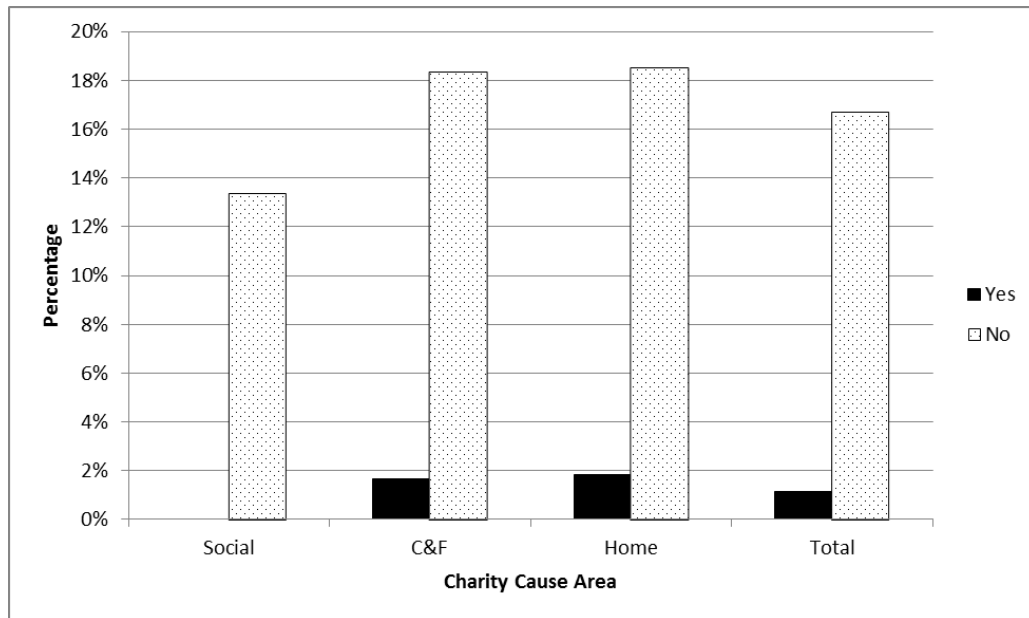
Family Services charities and 82 percent of Homeless Services charities answering either “reasonably confident” or “completely confident.” The percentages of evaluators having little confidence were similar across the three categories.

Figure 6.17: Percentage Breakdown of Element Three – Test Six: Does the organization benchmark its feedback data against other organizations? Binary Question (“Yes” responses to Test One)



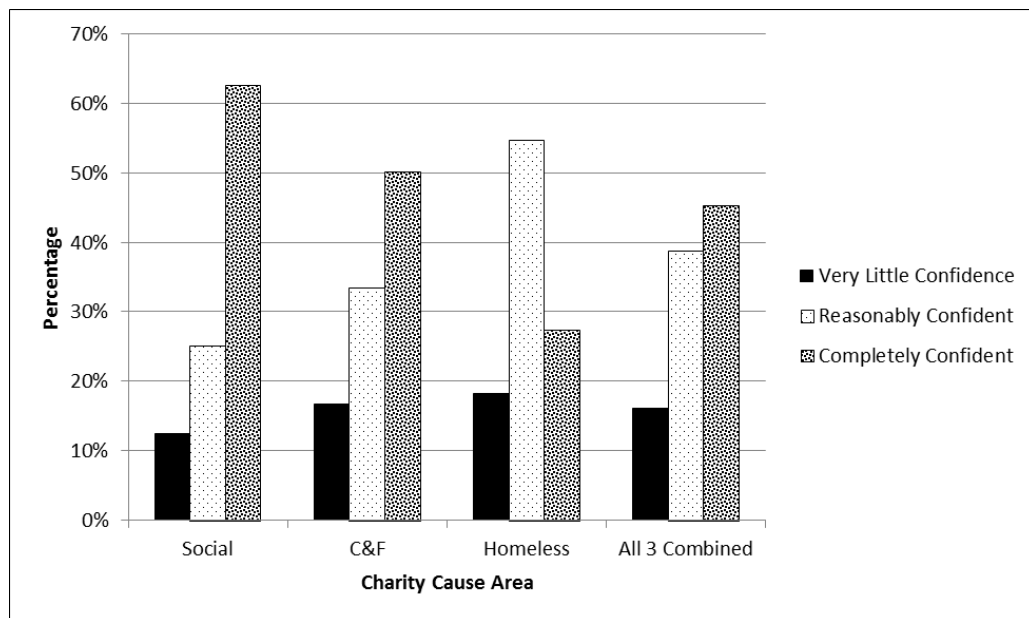
n=31

Figure 6.18: Percentage Breakdown of Element Three – Test Six: Does the organization benchmark its feedback data against other organizations? Binary Question (Total sample)



n=174

Figure 6.19: Percentage Breakdown of Element Three – Test Six: Does the organization benchmark its feedback data against other organizations? Researcher Confidence Level Responses



n=31

Qualitative: With Test Six, the only point of clarification necessary is regarding benchmarking. Some students were concerned whether or not it would be acceptable if an organization standardized its questions and compared results with affiliate organizations. For instance, many organizations look for certain standard outcomes because a larger “parent” seeks them. If an organization is a local chapter of United Way, for example, the “standardized questions” could come from United Way of America, which is not really an external organization. Is this “benchmarking”?

Furthermore, this question does not address whether organizations chosen as benchmarks should be comparable in scope or field to the reviewed charity. Nor does it specify the quality of the organizations selected for comparison, e.g., whether or not they are regarded as industry leaders. The concern the Capstone Group had is that the lack of detail about appropriate reference points could negatively incentivize charities to benchmark data against organizations less established, or facing particularly difficult challenges, making the charity look better than it otherwise might.

Observations from the Background Research Tool

With regard to Element Three, the Background Research Tool used nine groups of questions. They were developed chiefly on the basis of the conceptualization presented in *The 21st Century Potential of Constituency Voice*, one of whose authors has worked with Charity Navigator in developing CN 3.0 (Bonbright, Campbell, & Nguyen, 2009). As before, the Capstone Group used the BRT to analyze 89 charities. (See Appendix E.3 for the BRT research Frequency Tables.)

Research Question One: The first BRT question related to constituent voice asks for information about a charity's systems for collecting feedback: "Are there systems in place to gain input from primary recipients of service? Is this done through surveys, interviews, focus groups, or another way? Do any of the organization's outputs/outcomes use a client survey as a means of measurement? If so, do they publish those results?" According to *21st Century*, one of the purposes of feedback is capacity development, whereby organizations build the capability for constituents to give feedback (Bonbright, Campbell, & Nguyen, 2009, p. 16). The CNRT asked if nonprofits had "rigorously collected results." The BRT both addresses expands the concept to include examples of how feedback is collected. This allows for a non-binary answer that provides insight into what a charity is doing to collect constituent feedback.

For Research Question One, 15 of the 89 charities (17 percent) were found to have some kind of system in place to gain input from primary recipients of service. Because the BRT did not specify a "Yes" response for each method of collecting feedback, but was more open-ended, and some students simply answered "Yes" to the overall question, the Capstone Group does not have data about which method is most often utilized. In general, however, these 15 organizations seemed to have a commitment to gathering qualitative information from their clients, a proportion comparable to that obtained using the CNRT. Of the 15 charities, seven were Social Services charities, six were Children's and Family Services charities, and two were Homeless Services charities.

Research Question Two: The second BRT question for constituent voice asks for information about the utilization of feedback: "In the organization's annual report, newsletter, or website, did it highlight a donor, client or volunteer, using quotes or personal stories?" Since

the original CN test did not specify where to look for client feedback, this question was designed to guide the rater to look through a variety of sources to find general feedback. This question also expands the definition of constituent voice beyond that of the charity's beneficiary to include the experiences of donors and volunteers, utilizing constituent groups as defined in The Fund Raising School's Principles and Techniques of Fundraising (The Fund Raising School, 2011). However, the Capstone Group acknowledges that this question relates to the use of feedback for marketing, instead of its use to improve service delivery, as is intended with the concept of constituent voice.

For Research Question Two, 55 of the 89 charities (62 percent) were found to have a method of highlighting donor, client, or volunteer opinions. Of these charities, 21 came from Social Services, 15 were from Children's and Family Services, and 19 were Homeless Services groups.

Research Question Three: The third BRT question related to constituent voice asks for information about the scope and breadth of constituent voice: "Does the charity report how many people it engages in the community (including board members and volunteers)? Does it make a commitment to growth in this area?" One of the main themes of Keystone Accountability's framework is for constituent empowerment and engagement from the bottom-up as the most needed and relevant argument for constituent voice (Bonbright, Campbell, & Nguyen, 2009). Going beyond clients, this question addresses other constituencies as well and tries to see how much a charity values broader engagement.

For Research Question Three, 62 of the 89 charities (70 percent) were able to indicate how many members of the community were involved with the organization. These charities

were also seen as making a commitment to increasing community involvement. Of them, 21 came from Homeless Services and from Children's and Family Services, while 20 were from Social Services.

Research Question Four: The fourth BRT question related to constituent voice asks for information about the analysis and utilization of constituent voice: "Does the charity share the results of any surveys, interviews or focus groups with the public?" The purpose of this question is to connect feedback with results reporting. One of the values in *21st Century* is that feedback needs to be connected to results reporting as it shows interested parties performance measures and informs service and funding decisions (Bonbright, Campbell, & Nguyen, 2009, p. 34). Furthermore, by publicly reporting feedback, accountability to the general public is improved.

For Research Question Four, 20 of the 89 charities (23 percent) shared their survey and interview information with the public. Of these charities, five were from Social Services, 11 were from Children's and Family Services, and four came from Homeless Services.

Research Question Five: The fifth BRT question for constituent voice asks for information about the access constituents have for voicing their views: "Does the organization have an easy way of sending in comments, messages? Are feedback and comments encouraged in the charity's materials, website and publication? Do public documents state core commitment to empowerment, downward accountability, participation, being member-led, and other related concepts?" The purpose of this question is to measure the capacity charities have developed for obtaining feedback. Organizations that encourage engagement

give constituents a greater stake in the organization. This question is also predicated on the view that feedback efforts will be more sustained if cultivated from the bottom-up.

For Research Question Five, 30 of the 89 charities (34 percent) were found to either state core commitments to empowerment, downward accountability, and other related concepts, or encourage feedback and have ways for website visitors to easily make comments. Due to the double-barreled questions and binary nature of the BRT, the Capstone Group cannot determine the most common practices for sending in comments and messages. Of these charities, 16 came from Social Services, 11 from Children's and Family Services, and three from Homeless Services.

Research Question Six: The sixth BRT question related to constituent voice also asks for information about opportunities for constituent voice: "Does the charity list multiple ways to contact management or staff?" The purpose of this question is to ascertain the extent to which a charity is committed to collecting feedback from visitors to its website. The more contact information a charity provides, the more it shows it values constituent feedback because it is accessible to its primary constituents.

For Research Question Six, 69 of the 89 charities (78 percent) listed multiple ways to contact management, staff, or board members. Of these charities, 23 were from Social Services, 23 were from Children's and Family Services, and 23 were from Homeless Services.

Research Question Seven: The seventh BRT question for constituent voice asks for information about the manner of collecting constituent voice: "Does the charity host public meetings of any kind, board, staff, or other?" This question is designed to determine whether this mode of public input is available to constituents. One of the ways *21st Century* says

organizations can improve is to publicly report results back to constituents (Bonbright, Campbell, & Nguyen, 2009, p. 34). A way to do this is through meetings in the community, where a two-way transfer of information may take place. Thus, having public meetings can be a vehicle for accountability to the community, while simultaneously allowing for feedback collection as well.

For Research Question Seven, 17 of the 89 charities (19 percent) hosted public meetings. Due to different ways students recorded information about them, the Capstone Group cannot determine the most common forms of public meetings held by these charities. Among these charities, six were from Social Services, six from Children's and Family Services, and five from Homeless Services.

Research Question Eight: The eighth BRT question for constituent voice asks for information about the how constituent voice connects to organizational goals: "Does the charity have any goals or objectives related to the quality of service they provide or client satisfaction (i.e. - timely, good quality or friendly service)?" Even if a nonprofit has definite, qualitative goals, these may be difficult to quantify. An alternative way of measuring such outcomes would be through using results of constituent feedback. If a charity claims to have high quality or timely service, feedback may confirm or deny it.

For Research Question Eight, 40 of the 89 charities (45 percent) were found to have goals or objectives related to the quality of service they provide. Due to inconsistencies in recording, the Capstone Group cannot determine the most common goals represented by these charities. Of these charities, 14 were from Social Services, 14 were from Children's and Family Services, and 12 were from Homeless Services.

Research Question Nine: The last BRT question for constituent voice asks for information about the analysis and utilization of constituent voice: “Do any of the organization’s outputs/outcomes incorporate feedback as a means of measuring success (i.e. 42 of 60 clients surveyed said...)?” This question gives organizations the opportunity to collect and report feedback data as a part of outcome measurement. It also speaks to the organization’s ability to use feedback to improve program performance.

With Research Question Nine, eight of the 89 charities (9 percent) were found to incorporate feedback as a measurement of progress regarding organizational goals. Of these charities, one came from Social Services, five were from Children’s and Family Services, and two were from Homeless Services.

Phone Survey Analysis for Element Three

The Capstone Group conducted supplemental phone interviews for 35 charities after using the online CNRT in order to better understand the extent to which charities collected data for Element Three, but did not show it on their websites. From examining the consistencies or inconsistencies between information online and information housed internally at charities, the Capstone Group could determine what information charities have available but are not publicly reporting. Although the CNRT is designed to measure public reporting of the information related to constituent voice, insight as to whether charities have the relevant information available internally, and their reasons for not publicly reporting it, should be useful.

In the phone interviews, 22 charities mentioned the use of Element Three data collection techniques that were not appearing on their websites. This figure would increase the finding of the CNRT that (31 charities collected feedback by about 70 percent. This large

increase indicates that the CNRT is not capturing all of charities' current practices in collecting information related to constituent voice, though it successfully captures the public reporting of that information, which is its intent. The Capstone Group feels that rating the public reporting of this concept will incentivize charities to publish relevant information on their websites.

To better understand the difference between data collection and publishing results, the Capstone Group asked charities to disclose current limitations in their ability to publish collected data. The most repeated concern of the charities was privacy (14 charities, or 40 percent); publishing feedback from constituents, the organizations that withheld it believed, had the potential to intrude into beneficiaries' private opinions and outcomes. Even if it were aggregated, such information can be sensitive as it can speak to the character, life choices, and vulnerabilities of the beneficiaries.

The second most common set of concerns was with cost and technological capacity (for both, eight charities, or 23 percent). Several charities noted that their websites or staff knowledge for maintaining their websites simply did not allow for presenting feedback information. Charities also cited limitations in staff time (seven charities, or 20 percent), which may speak directly to the next highest ranked concern: that their programs and feedback data are too complicated to translate into a public-friendly presentation (six charities, or 17 percent). Finally, a few charities said that there simply is not enough public demand for them to go through the trouble of posting the information (three charities, or nine percent).

In addition to discovering why some charities do not publish constituent voice data on their websites when they do collect the data, the Capstone Group also asked if the charities saw value in the idea of CN's use of Element Three, regardless of whether they currently collect

relevant data or not. Many respondents (12 charities, or 34 percent) felt that there was value in the constituent voice framework, but a third of these (four charities of the 12) insisted that some information needs to remain internal. The next most frequent concern was that the new CN framework creates the potential for complicated and difficult comparisons between very different organizations with diverse environments (seven charities, or 20 percent). Finally, charities thought that Element Three's concept was not well-suited for collecting quantitative data, or they were uncertain how to collect quantitative data on qualitative feedback (five charities, or 14 percent).

The Capstone Group's findings from the phone interviews indicate that the CNRT should undergo some changes in how it tries to ascertain how charities collect and report on constituent voice. CN may wish to give greater consideration to the charities' concerns about comparing constituent voice practices across the charitable sector when the diversity and vulnerabilities of each charity's constituents may necessitate very different approaches to and methods of collecting and using feedback.

Chapter Seven

Element Four -- Alignment of Mission and Resources

The goal of the Element Four: Alignment of Mission and Resources is to determine the extent to which a charity's allocation of resources aligns with its stated mission and goals. This Element assesses whether or not charities are being responsible stewards of their resources by applying them to where they are seeking to have the most impact. Because the SPEA Capstone Group began testing the Charity Navigator Research Tool before this fourth element was fully developed, it was unable to apply it fully. However, the CNRT included questions related to the Compliance Footprint concept in Element One (Logic, Results, and Measures) and Element Two (Independent Evaluations, Reporting Standards and Certification Mechanisms), which is very similar conceptually to Element Four. In answering the Compliance Footprint questions, students were asked to indicate – using a percentile scale – what proportion of a charity's programs reflect the concepts discussed in Elements One and Two, specifically results-measurement and independent evaluations or certifications. Thus, for purposes of this test of CN 3.0, the SPEA Capstone Group used the Compliance Footprint questions as a surrogate for Element Four, since they both aim to measure the extent to which a charity's resources align with its mission and goals. The following section of the report presents and discusses the Compliance Footprint findings from both Elements and considers how Charity Navigator can use these results for developing criteria for Element Four.

Compliance Footprint – Element One

Each of Element One’s four tests asked students to assign a percentage score for how many of a charity’s programs report using logic models, results, or data-collection methods. According to the CN Research Guidelines, the Compliance Footprint “tells us how much of the charities’ mission is covered by the statements that passed the four tests” (Charity Navigator, 2012, p. 4).

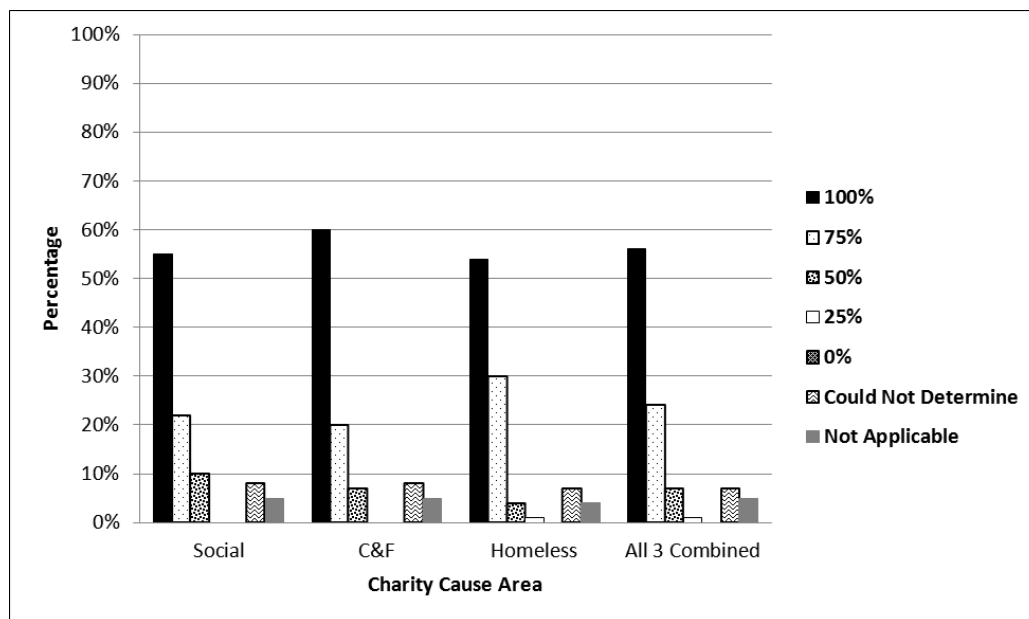
Students filled out a total of 174 surveys for the Charity Navigator Research Tool. Since these surveys covered charities in three cause areas, the percentages provided below *for all three cause areas combined* (Social Services, Children’s and Family Services, and Homeless Services) are out of a total of 174. The percentages provided for Social Services and Children’s and Family Services organizations are both out of a total of 60 student-completed online surveys, and the percentages provided for Homeless Services organizations are out of a total of 54 student-completed online surveys. The Compliance Footprint and Inter-Rater Reliability results for the Logic, Results, and Measures Element’s four tests are summarized and are reported in the Frequency Tables (See Appendix C.) Student responses to the CNRT’s open-ended questions were analyzed in order to understand the challenges that they confronted. The issues relevant to applying the Compliance Footprint in Element One are thus identified and explained.

Test One: Is the Causal Logic Plausible?

Quantitative: Figure 7.1 shows that for the test’s Compliance Footprint section, students selected the “100%” and “75%” of programs are compliant as answer choices the majority of the time, while students selected the “50%,” “25%,” “0%,” “Could Not Determine,”

and “Not Applicable” answers each less than 12 percent of the time. These findings held both for all cause areas combined and for each area individually. This means that a large majority of the charities’ programs are compliant and showed some plausible causal logic.

Figure 7.1: Percentage Breakdown of Element One – Test One: Is the Causal Logic Plausible? Compliance Footprint Responses



n=174

Qualitative: Students expressed concerns regarding the Compliance Footprint aspect of Test One because there were instances in which they wanted to choose a value that landed between the available answer options, which left students unsure about whether to round up or down.

Inter-Rater Reliability: Figure 7.2 shows that, both for all three cause areas combined and for each area individually, more student-pairs had different answers for the Compliance Footprint question than had identical answers in Test One. In situations where both students provided a numerical response and there was a difference in how the two students answered,

the average difference between the two responses was one level, or 25 percentage points, as shown in Figure 7.3. Thus, even though more student-pairs had different answers than had identical answers, the average discrepancy between the two responses generally amounted to one-level difference (a 25 percentage point increment) on the scale used by the CNRT. This finding held both for all three cause areas combined and each area individually, as reflected in Figure 7.3.

Figure 7.4 shows the number of instances in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”) or in which both students provided non-numerical responses compared to the total number of discordant responses for the Compliance Footprint question. These kinds of discrepancies in responses occurred slightly more often in surveys of Social Services organizations and Children’s and Family Services organizations (7 times for each) than they did in surveys for Homeless Services organizations (5 times). The results also indicate that this kind of discrepancy was observed in less than half of discordant pairs for Test One; only in 19 out of the 53 total discordant pairs did one student provide a numerical response and the other a non-numerical response or both students choose non-numerical responses. This indicates that disagreements among students, though prevalent, were relatively small. Only a few pairs of students drastically disagreed by one choosing a numeric answer and the other, a non-numeric answer of “Not Applicable” or “Could Not Determine.”

Figure 7.2: Percentage Breakdown of Inter-Rater Reliability for Element One – Test One: Is the Causal Logic Plausible? Compliance Footprint Responses

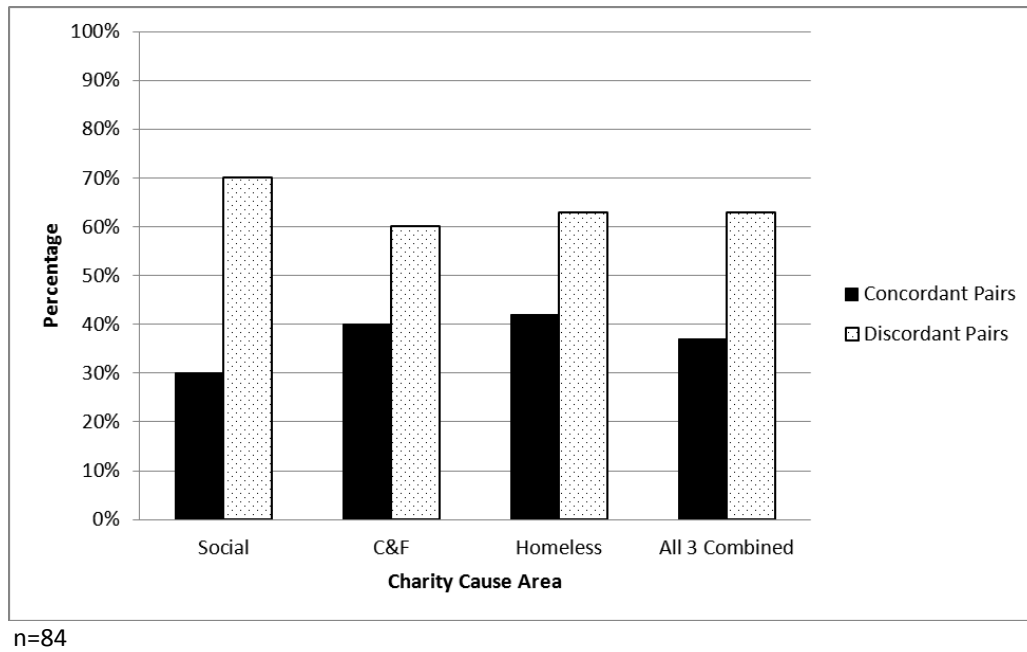


Figure 7.3: Average Point Spread in Compliance Footprint Responses of Discordant Pairs for Element One – Test One: Is the Causal Logic Plausible? (on a scale from 0 to 100)

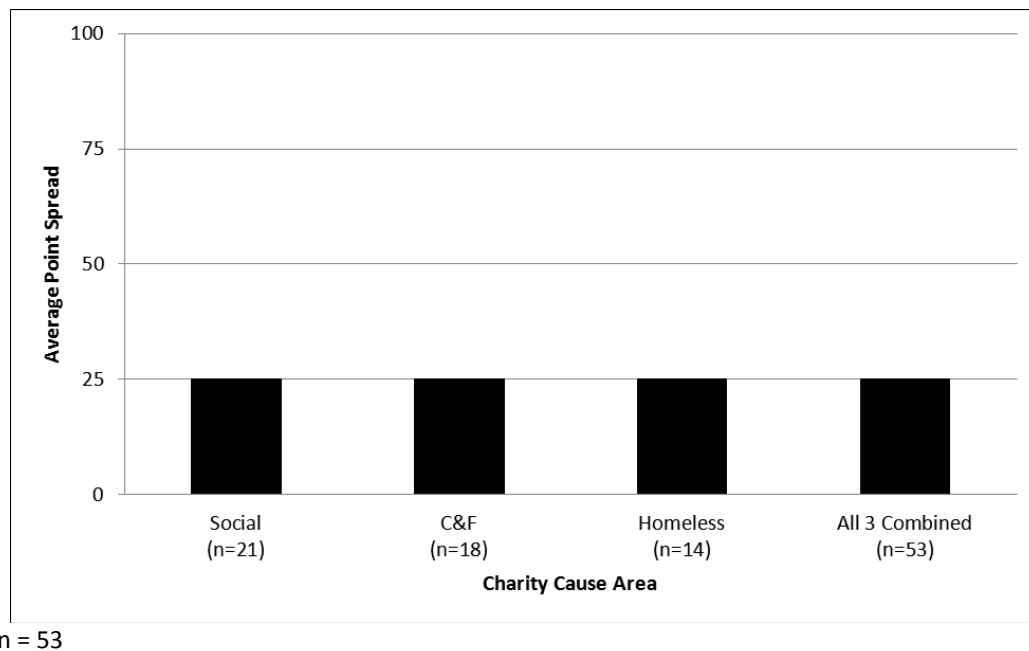
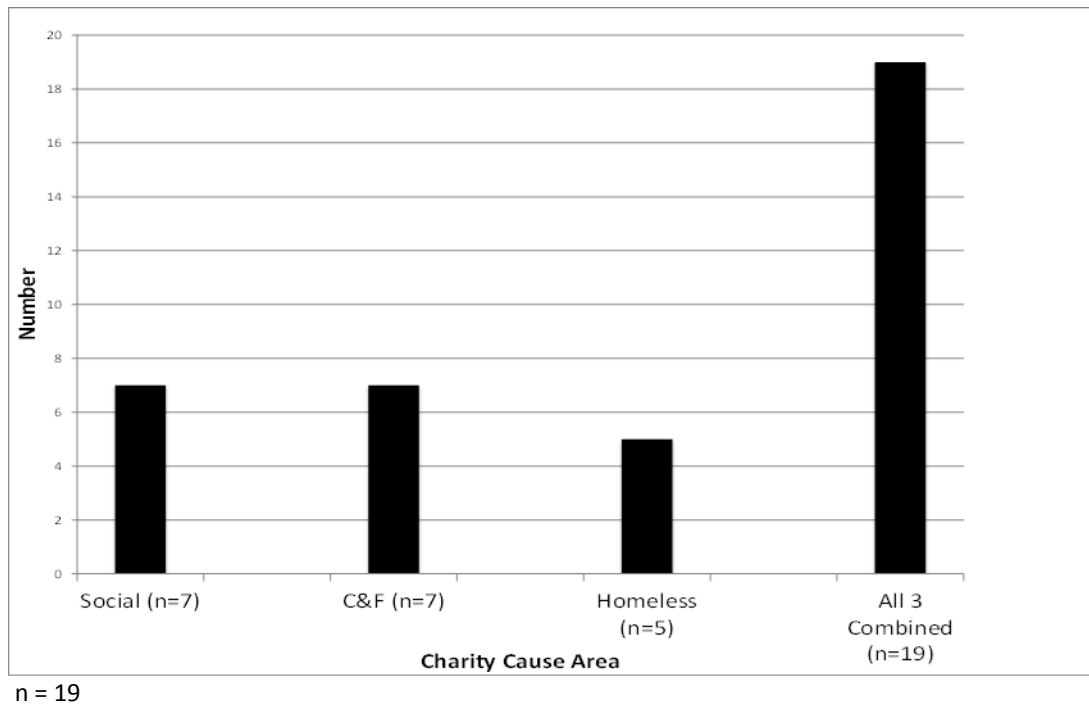


Figure 7.4: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or 2 Non-Numerical Scores in Compliance Footprint Question for Element One – Test One: Is the Causal Logic Plausible?

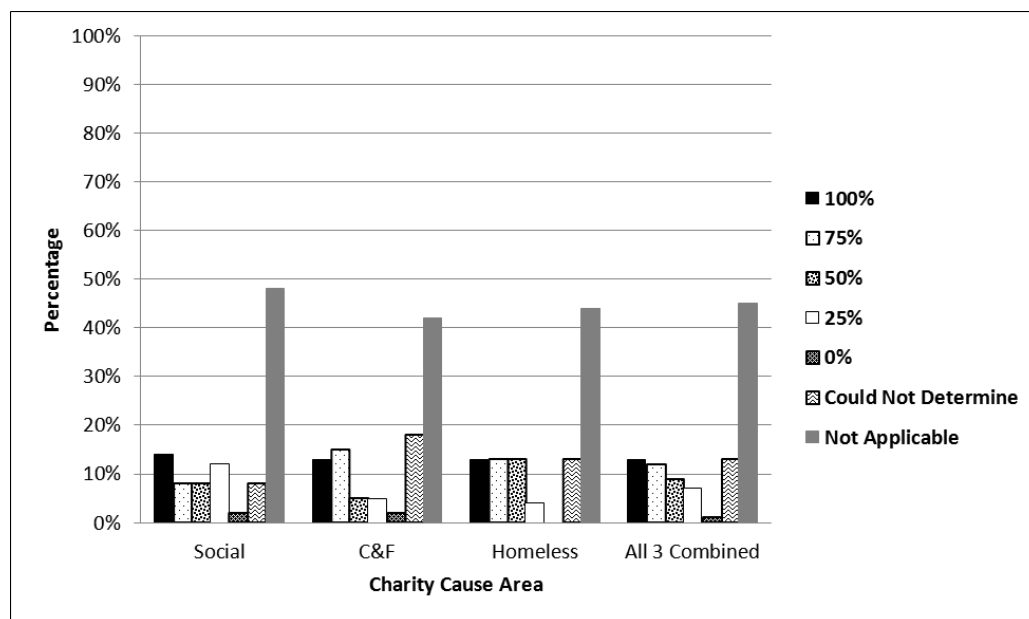


Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)?

Quantitative: As Figure 7.5 shows, for all three cause areas combined in the test’s Compliance Footprint section, students answered “Not Applicable” most often, at more than 40 percent of the time, while “100%,” “75%,” and “Could Not Determine” were all chosen slightly more than 10 percent of the time, and “50%,” “25%,” and “0%” less than 10 percent of the time. The distribution of answers for the individual cause areas was more varied than it was for all cause areas in the aggregate. However, as with all three cause areas combined, “Not Applicable” was chosen the most often for each area individually, as reflected in Figure 7.5.

This indicates that either students could not find the information on the website to determine dosage, or that students could not determine from the dosage information provided to which percentage of a charity's programs the published dosage pertained.

Figure 7.5: Percentage Breakdown of Element One – Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)? Compliance Footprint Responses



n=174

Qualitative: Similarly to Test One, students had concerns about which answer to select when the Compliance Footprint percentage was closer to an amount between the answer options. Students also said that having a clearer example of how to calculate the Compliance Footprint in the Charity Navigator Research Guidelines would be helpful. There was also a great deal of confusion about how to respond to this question when an organization had a multitude of programs. For example, Jewish Vocational Services has nine different programs, such as *Career Services and Training, Disability and Assessment, Community Services, a Scholarship Program*, and others. Each program has its own section within the website, but whether or not

information regarding dosage was provided varied greatly, which left students unsure about how to calculate the Compliance Footprint for the organization as a whole. For this test, both for all three cause areas combined and each area individually, “Could Not Determine” and “Not Applicable” combined for almost 60 percent of responses, which may reflect these uncertainties.

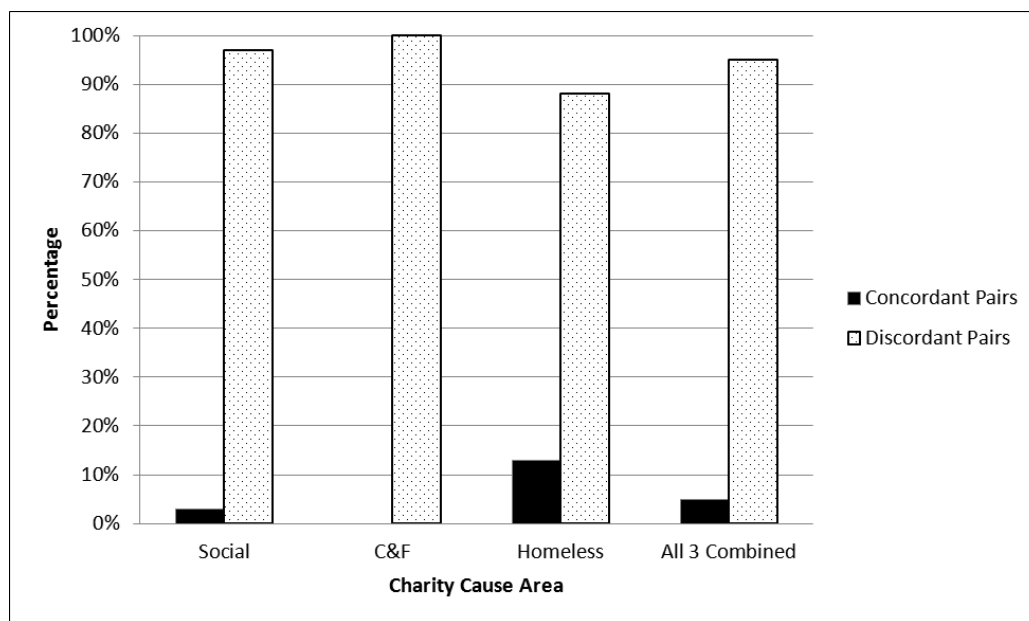
Inter-Rater Reliability: Figure 7.6 shows that, both for all three cause areas combined and each area individually, the vast majority of student-pairs did not have identical answers for the Compliance Footprint question in Test Two, or misunderstood the concept of “dosage” in this Test. In situations where both students provided a numerical response and there was a difference in how the two students answered, the average difference between the two responses equaled two levels in the scale (50 percent of programs) for all three cause areas combined, Social Services organizations, and Homeless organizations, while the average difference equaled one level (25 percent) for Children’s and Family Services organizations, as seen in Figure 7.7. This finding indicates that there was not only a large number of discordant pairs, but the average discrepancy between the responses was, on the whole, more substantial than it was in Test One.

Figure 7.8 shows the number of instances in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”) or in which both students provided non-numerical responses, out of the total number of discordant responses for the Compliance Footprint question. According to this Figure, these kinds of discrepancies in responses occurred much more often in surveys for Children’s and Family Services organizations (27 times) and Social Services organizations (25

times) than they did in surveys for Homeless Services organizations (6 times). They were also observed in the majority of discordant pairs for Test Two because in 58 out of the 80 total discordant pairs, either one student provided a numerical response and the other a non-numerical response or both students chose non-numerical responses.

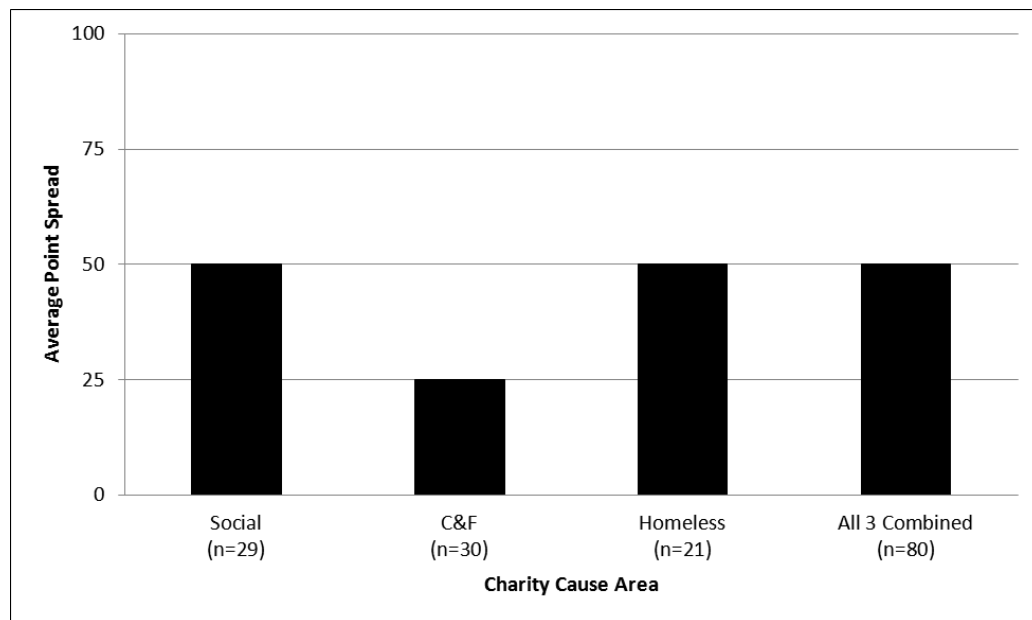
The findings of this test show student confusion over applying the concept of Compliance Footprint to a question about dosages or difficulty in finding the information that would have led to similar answers. A more sensitive scale than one divided into 25-percent increments could reduce the difficulty in choosing between large levels. However, the fact that almost all student pairs disagreed shows the difficulty of determining a Compliance Footprint in this area of results reporting.

Figure 7.6: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)? Compliance Footprint Responses



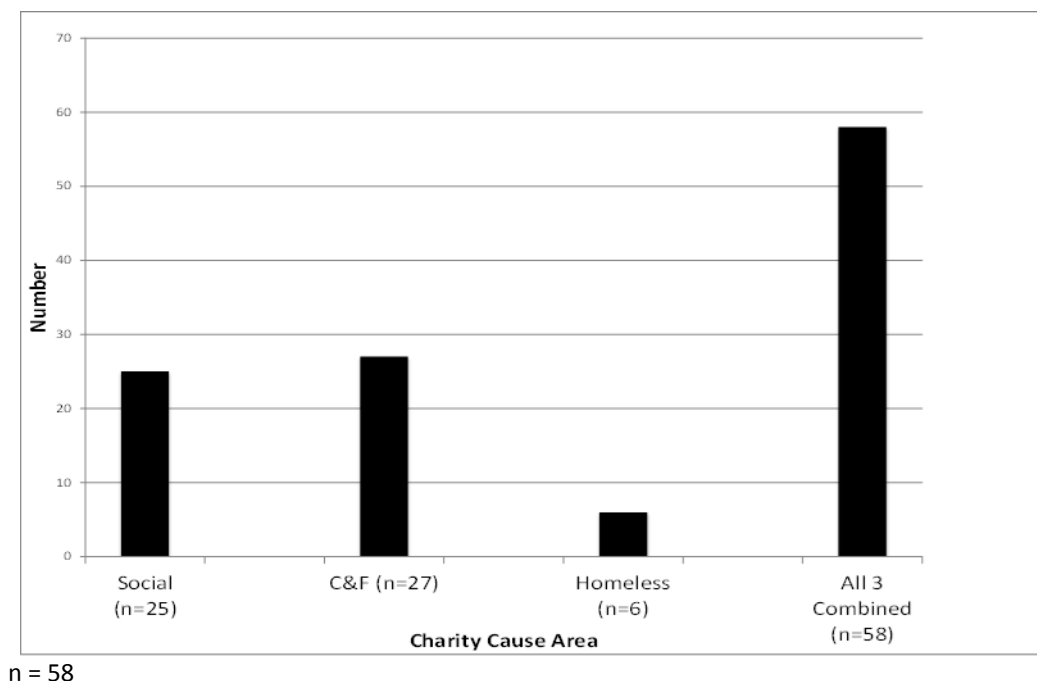
n=84

Figure 7.7: Average Percentage Point Spread in Compliance Footprint Responses of Discordant Pairs for Element One – Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)? (on a 25-percentage scale from 0 to 100)



n = 80

Figure 7.8: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or 2 Non-Numerical Scores in Compliance Footprint Question for Element One – Test Two: Is there some indication of how much of the action is required to produce what effects (i.e., dosage)?



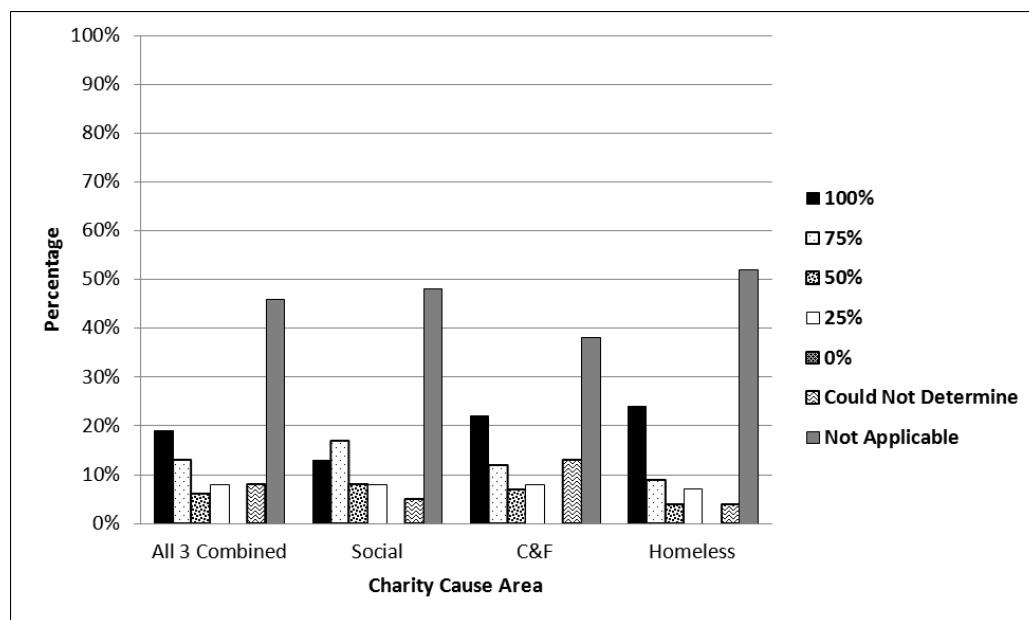
Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)?

Quantitative: Figure 7.9 shows that, for all three cause areas combined in the test's Compliance Footprint section, students answered "Not Applicable" the most often, at more than 45 percent of the time, while the "100%" and "75%" answers were selected between 10 percent and 20 percent of the time, and "50%," "25%," "0%," and "Could Not Determine" less than 10 percent of the time. The findings also show that none of the students chose "0%." As with Test Two, the distribution of answers for the individual cause areas was more varied than

it was for all cause areas in the aggregate. However, as with all three cause areas combined, “Not Applicable” was chosen the most often, as Figure 7.9 reflects.

This reason “Not Applicable” was chosen almost 50 percent of the time is that the largest percentage of charities did not reasonable evidence as a foundation for their programs. However, when reasonable evidence was available, almost 30 percent of charities had all or almost all of their programs compliant.

Figure 7.9: Percentage Breakdown of Element One – Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)? Compliance Footprint Responses



n=174

Qualitative: Students’ concerns about the Compliance Footprint in Test Three were similar to those for the previous tests, including how to know what response to choose when the most appropriate value seemed to equal a number in between the answer choices, not having more specific examples for making the Compliance Footprint calculation, and whether the question should be applied to the organization as a whole or to its individual programs.

This confusion most likely contributed to the fact that for all three cause areas together, the “Could Not Determine” and “Not Applicable” answers combined for almost 55 percent of responses and that “Not Applicable” was chosen the most often for each area individually.

Inter-Rater Reliability: Similar to Test Two, the vast majority of student-pairs had discordant answers for the Compliance Footprint question in Test Three, for both all three cause areas combined and each area individually, as seen in Figure 7.10, although possibly, the student answers reflected a misunderstanding of the meaning of “reasonable evidence” in this Test. For discordant pairs in which both students provided a numerical response, the average difference between the two responses equaled one level of the scale (25 percentage points) for all three cause areas combined, Social Services organizations and Homeless Services organizations, while the average difference equaled two levels of the scale (50 percentage points) for Children’s and Family Services organizations, as Figure 7.11 shows. This finding indicates that although there were a large number of discordant pairs, the average discrepancy between the responses was, on the whole, smaller than it was in Test Two.

Figure 7.12 shows the number of instances in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”), or in which both students provided non-numerical responses, out of the total number of discordant responses for the Compliance Footprint question. These kinds of discrepancies in responses occurred slightly more often in surveys for Social Services organizations and Children’s and Family Services organizations (22 times for each) than they did in surveys for Homeless Services organizations (17 times). But that may simply be due to the fact that there were more total discordant pairs observed for Social Services and Children’s and

Family Services organizations than for Homeless Services organizations. Figure 7.12 also indicates that this kind of discrepancy was observed in almost all discordant pairs for Test Three; in 61 out of the 79 total discordant pairs, either one student provided a numerical response and the other a non-numerical response or both students chose non-numerical responses. This may indicate problems in wording of the test, too broad of a scale, or a lack of charity information or student understanding, resulting in disagreement among pairs.

Figure 7.10: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)? Compliance Footprint Responses

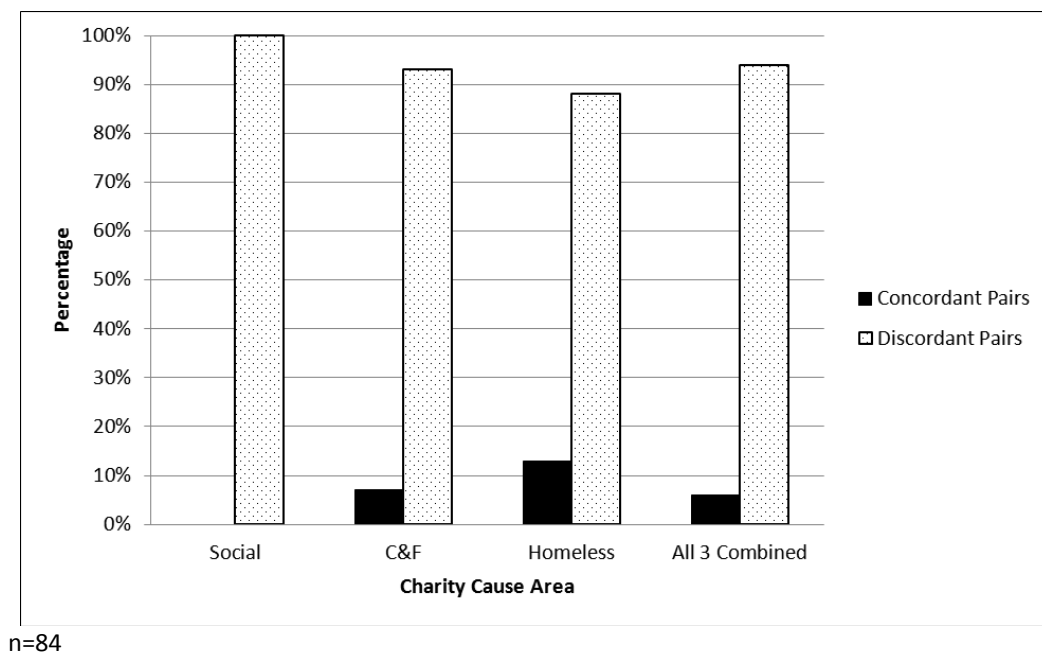
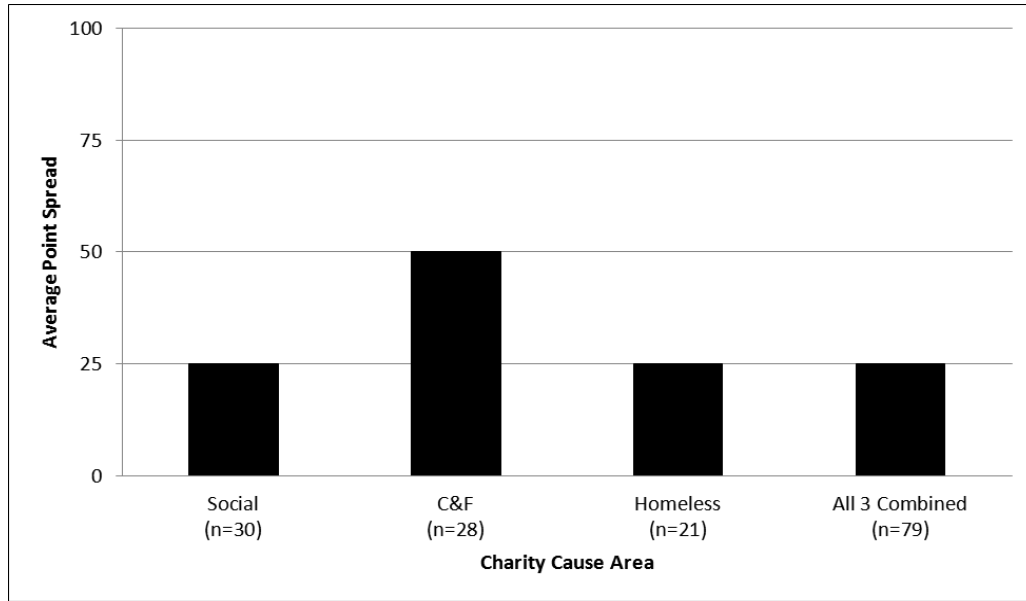
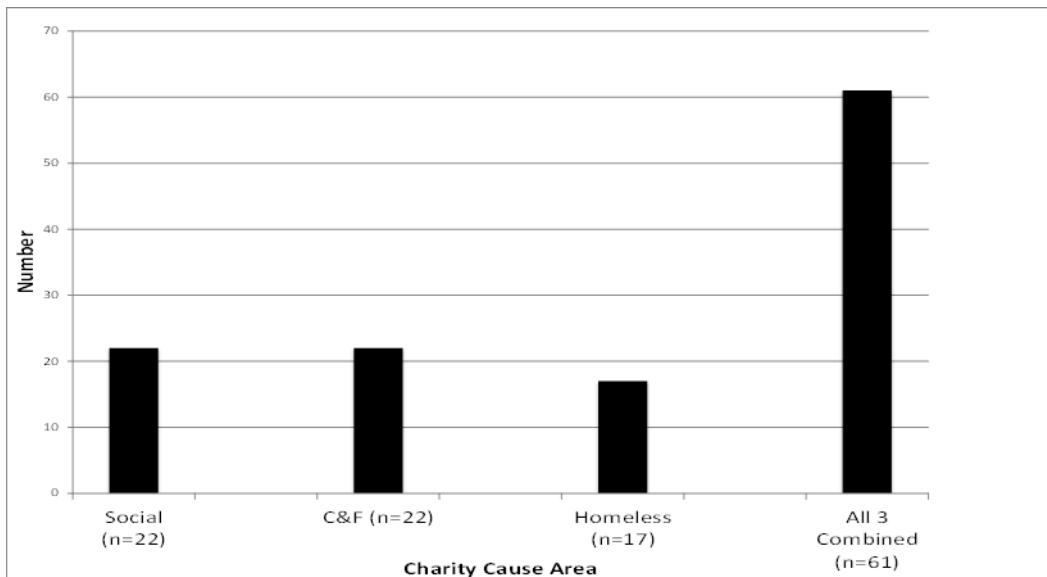


Figure 7.11: Average Percentage Point Spread in Compliance Footprint Responses of Discordant Pairs for Element One – Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)? (on a 25-percent scale from 0 to 100)



n = 79

Figure 7.12: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or 2 Non-Numerical Scores in Compliance Footprint Question for Element One – Test Three: Is there some indication that the logical inference in (1) is based on reasonable evidence (i.e., references to other programs that have demonstrated success with this logic)?

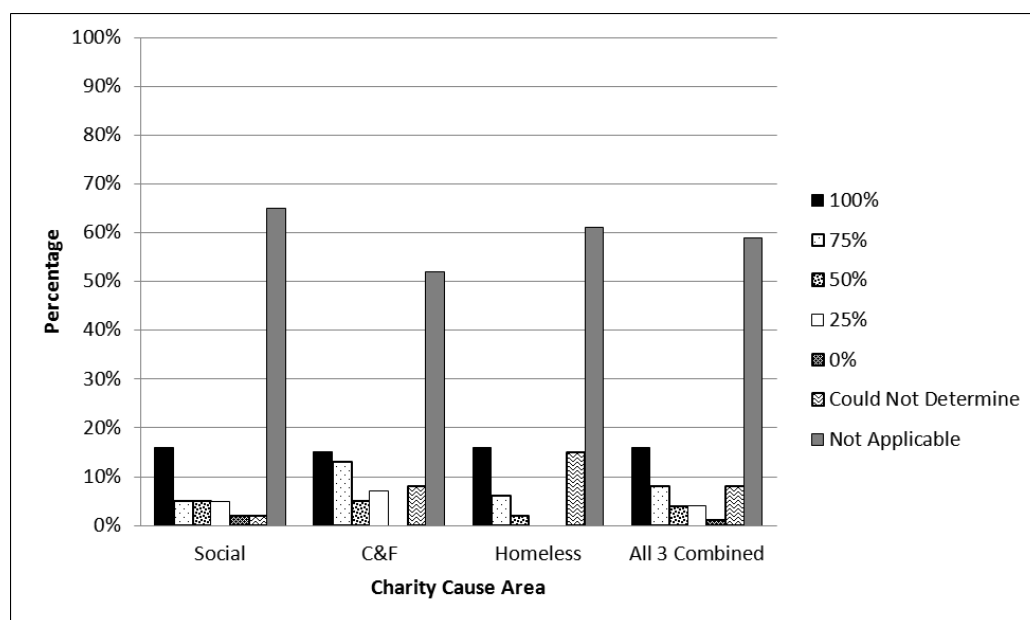


n = 79

Test Four: Are there specified measures (indicators) to be collected and a plan to do so?

Quantitative: Figure 7.13 shows that, for all three cause areas combined in the test’s Compliance Footprint section, students answered “Not Applicable” most often, at more than 55 percent of the time, while the “100%” answer choice was selected slightly less than 20 percent of the time and the “75%,” “50%,” “25%,” “0%,” and “Could Not Determine” choices less than 10 percent of the time. As with Tests Two and Three, the distribution of answers for the individual cause areas was more varied than it was for all cause areas in the aggregate. However, as with all three cause areas combined, Figure 7.13 illustrates “Not Applicable” was chosen the most often, followed by “100%.”

Figure 7.13: Percentage Breakdown of Element One – Test Four: Are there specified measures (indicators) to be collected and a plan to do so? Compliance Footprint Responses



n=174

Qualitative: As with the three previous tests, students expressed uncertainty about answering the Compliance Footprint question for a variety of reasons, including that the intervals in answer choices were too broad, there were instances when it was not clear to what extent information about measures (indicators) to be collected and a plan to do so was applied to all of an organization’s programmatic elements, and the Research Guidelines did not offer a straightforward example of how to calculate the Compliance Footprint. The fact that “Could Not Determine” and “Not Applicable” combined for almost 70 percent of responses for all cause areas together and that students chose “Not Applicable” far more often than any of the other answers for each cause area individually can most likely be attributed to these uncertainties.

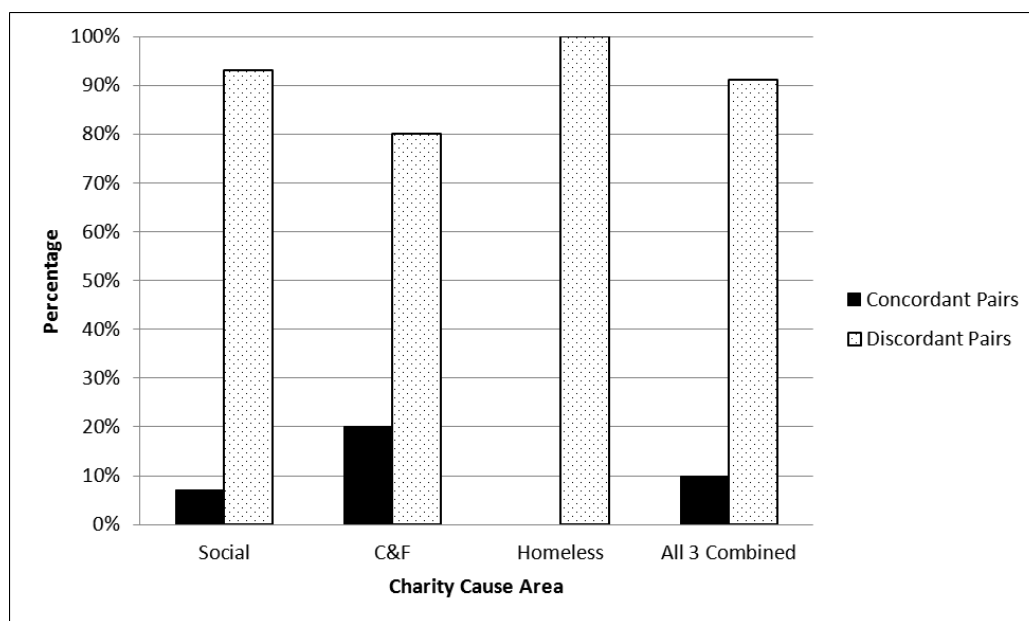
Inter-Rater Reliability: Similar to Tests Two and Three, the vast majority of student-pairs had discordant answers for the Compliance Footprint question in Test Four, for both all three cause areas combined and each area individually, as seen in Figure 7.14. In situations where both students provided a numerical response and there was a difference in how the two students answered, the average difference between the two responses equaled one level on the scale (25 percentage points) as shown in Figure 7.15. Thus, even though a large proportion of student-pairs had discordant answers, the average discrepancy between the two responses amounted to a small number. This finding held both for all three cause areas combined and each area individually.

Figure 7.16 shows the number of instances in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”), or in which both students provided non-numerical responses, out of the

total number of discordant responses for the Compliance Footprint question. These kinds of discrepancies in responses occurred most often in surveys of Social Services organizations (26 times), followed closely by surveys of Homeless Services organizations (22 times) and Children's and Family's Services organizations (21 times). In 69 out of the 76 total discordant pairs, either one student provided a numerical response and the other a non-numerical response or both students chose non-numerical responses.

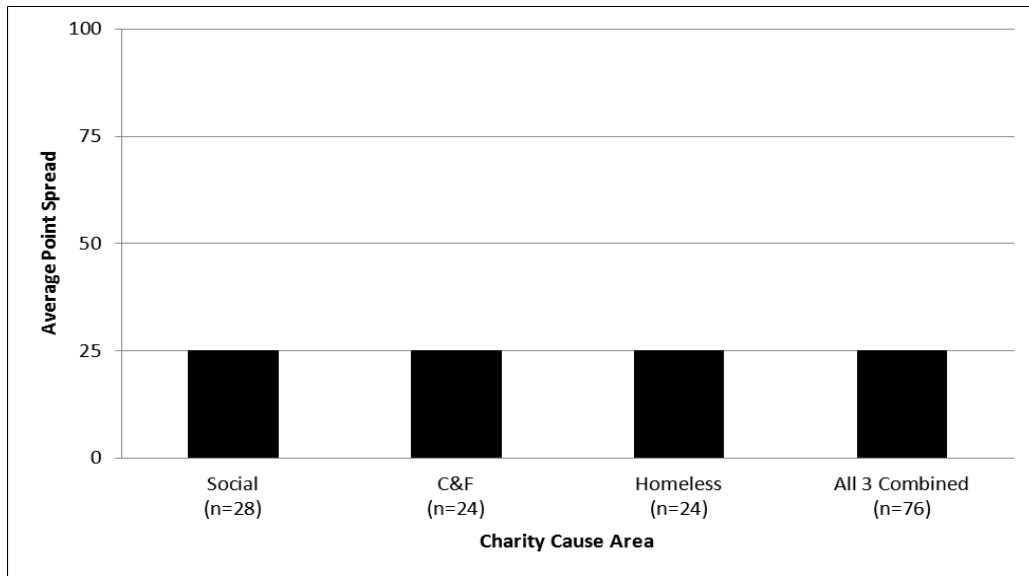
This means that measures or data collection plans for charity programs were either not found by both students or not published on the website at all. Other causes of the wide variation observed could be student difficulty with the scale intervals of 25 percentage points, unclear wording of the question, or a lack of understanding how to calculate compliance footprint on the part of students.

Figure 7.14: Percentage Breakdown of Inter-Rater Reliability for Element One – Test Four: Are there specified measures (indicators) to be collected and a plan to do so? Compliance Footprint Responses



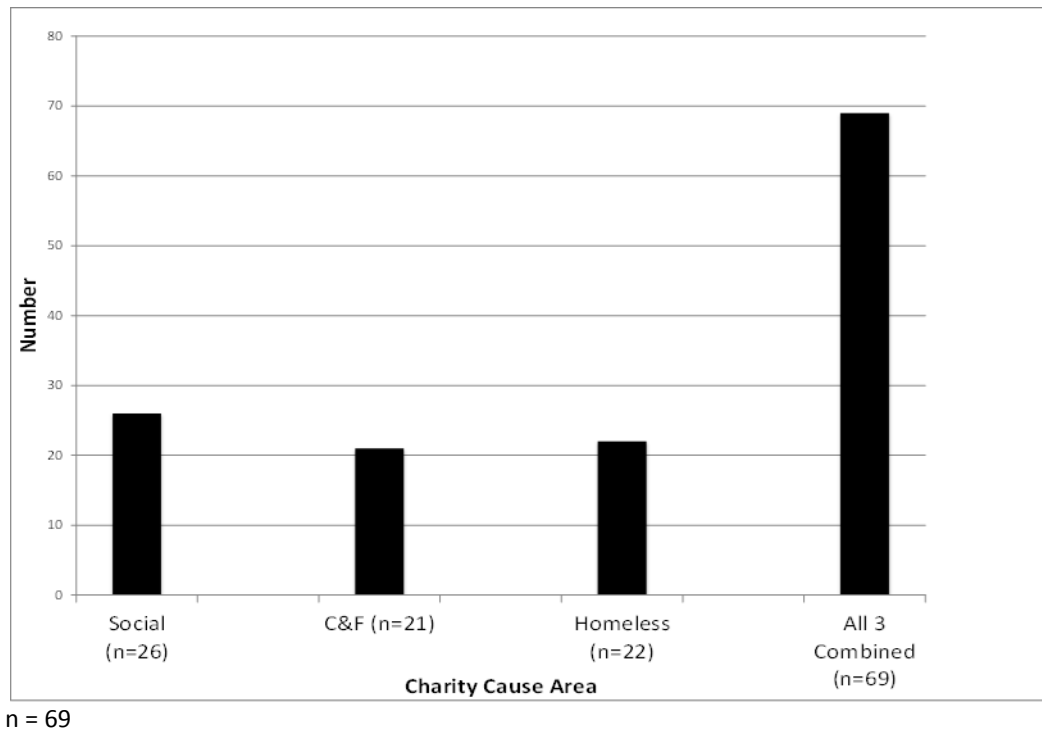
n=84

Figure 7.15: Average Percentage Point Spread in Compliance Footprint Responses of Discordant Pairs for Element One – Test Four: Are there specified measures (indicators) to be collected and a plan to do so? (on a 25-percent scale from 0 to 100)



n = 76

Figure 7.16: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or 2 Non-Numerical Scores in Compliance Footprint Question for Element One – Test Four: Are there specified measures (indicators) to be collected and a plan to do so?



Compliance Footprint - Element Two

Both of Element Two's tests asked students to assign a percentage score for Compliance Footprint following the binary questions. According to the CN Research Guidelines, the Compliance Footprint tells us "how much of the total work of the organization is covered by published evaluations" and to what extent "standards and certifications are applied to the whole of the organization" (Charity Navigator, 2012, pgs. 9, 10).

As in Element One, the number of student responses for all three cause areas combined is 174, while for Social Services and Children's and Family Services organizations, it is 60 each and for Homeless Services, 54. The Compliance Footprint results for Element Two's tests are summarized and are reported in the Element Two Frequency Tables. (See Appendices D.1 and

D.2.) Again, students' open-ended responses in using the Tool were analyzed and the issues relevant to the Compliance Footprint in Element Two are assessed.

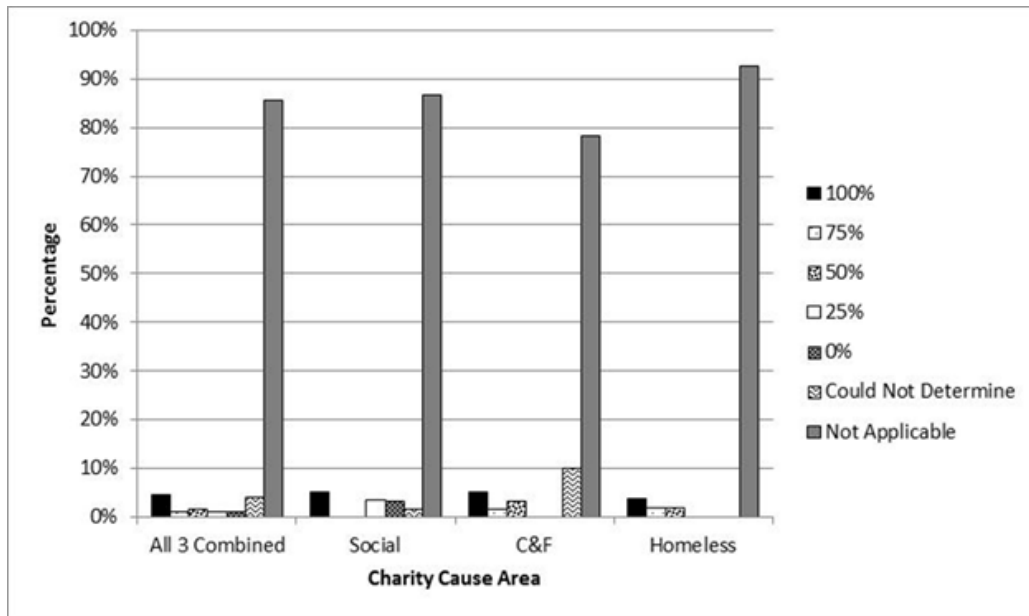
Test One: Independent Evaluations

This test examines whether the charity publishes an independent evaluation and whether the charity publishes progress reports on corrective actions undertaken following the evaluation. These questions were originally written by CN as "Published independent evaluations?" and "Progress reports on corrective actions undertaken following the evaluation?"

Quantitative: Figure 7.17 shows that for the test's Compliance Footprint section, students selected "Not Applicable" the majority of the time. Students selected the "100%," "75%," "50%," "25%," and "0%" answer choices no more than five percent of the time. Students chose "Could Not Determine" ten percent of the time or less. These findings held both for all cause areas combined and each area individually.

Overall, even though the vast majority of responses indicate that the Compliance Footprint was not applicable, the majority of the 18 responses which found evidence of independent evaluations had a fairly high average Compliance Footprint (77 percent), indicating that, on average, the charities that have independent evaluations cover more than three-fourths of their programs with them. These data suggest that charities that are already convinced of the value of reporting independent evaluations information are indeed using them to broadly cover their programs, but the majority of reviewed charities are not employing or reporting on such evaluations.

Figure 7.17: Percentage Breakdown of Element Two – Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?” Compliance Footprint Responses



n=174

Qualitative: Students expressed concerns regarding the Compliance Footprint aspect of Test One because many charities offered unclear information on their websites about their evaluations. There is no standard method of sharing this information, which made it difficult to find and assess.

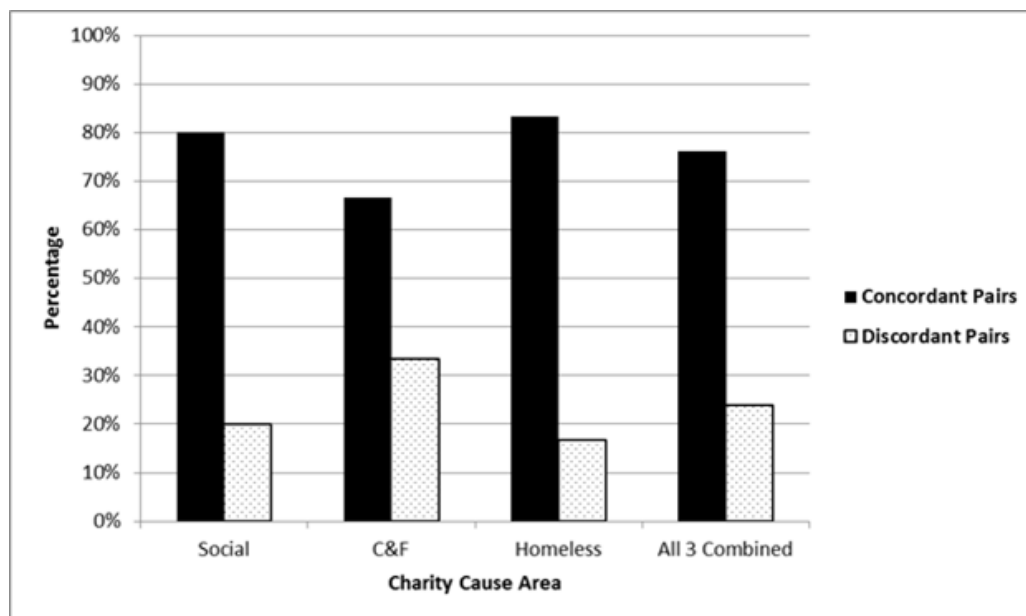
Inter-Rater Reliability: Figure 7.18 shows that the percentage of concordant pairs is greater than the number of discordant pairs for all three cause areas. In situations where both students in a pair provided numerical responses and there was a difference in how the two students answered, all of the pair differences were summed and the total difference was divided by the number of discordant pairs to produce the average percentage difference in student-pairs. Homeless Services organizations had the highest percentage of concordant pairs

(83 percent), with Social Services organizations close behind at 80 percent. Children’s and Family Services organizations had the lowest percentage of concordant pairs (67 percent).

Figure 7.19 shows that neither Social Services nor Homeless Services organizations had any charities for which both students provided a numerical answer. For both of these cause areas, the average spread for Compliance Footprint cannot be calculated. For Children’s and Family Services organizations, the average spread in Compliance Footprint is two levels of the scale (50 percentage points), as shown in Figure 7.19.

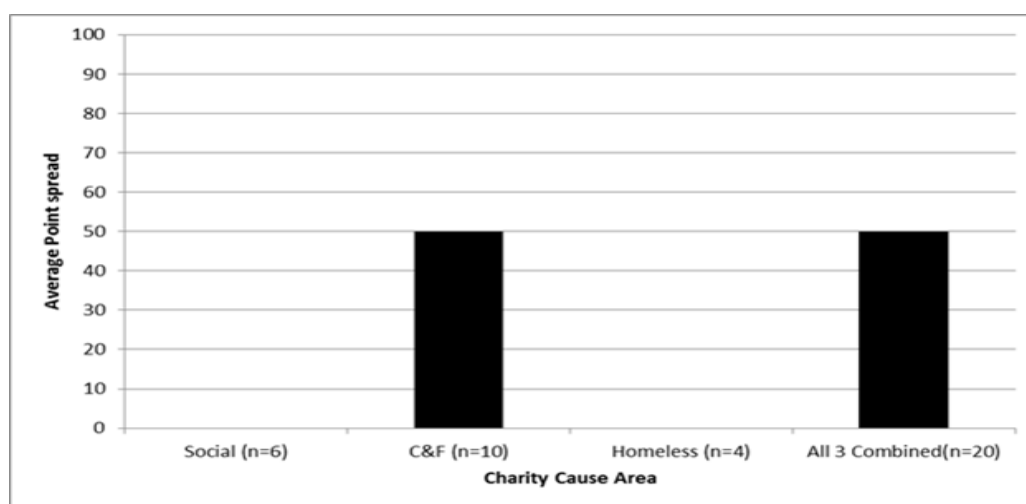
Figure 7.20 shows the number of instances in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”), or in which both students provided non-numerical responses, out of the total number of discordant responses for the Compliance Footprint question. These kinds of discrepancies in responses occurred slightly more often in surveys for Children’s and Family Services organizations (nine times for each) than they did in surveys for Social Services organizations (six times) and Homeless Services organizations (four times). They were observed for almost all of the discordant pairs for Test One because in 19 out of the 20 total discordant pairs, one student provided a numerical response and the other a non-numerical response, or both students choose non-numerical responses.

Figure 7.18: Percentage Breakdown of Inter-Rater Reliability for Element Two – Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?” Compliance Footprint Question Responses



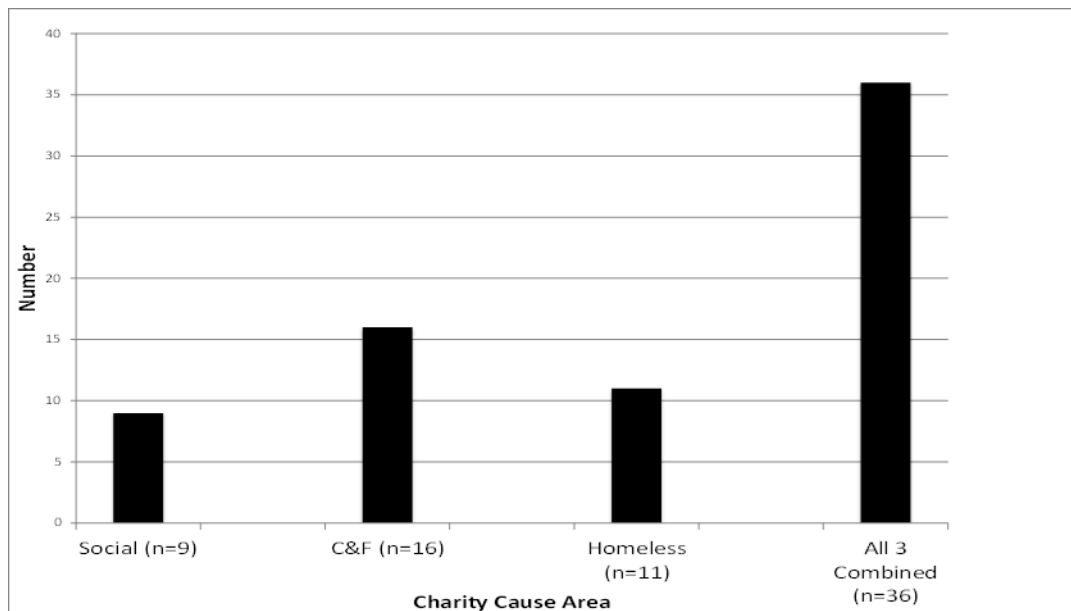
n=84

Figure 7.19: Average Percentage Point Spread in Compliance Footprint Responses of Discordant Pairs for Element Two – Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?” (on a scale from 0 to 100)



n = 20. Social and Homeless Services organizations had no discordant pairs in which both students provided numerical answers.

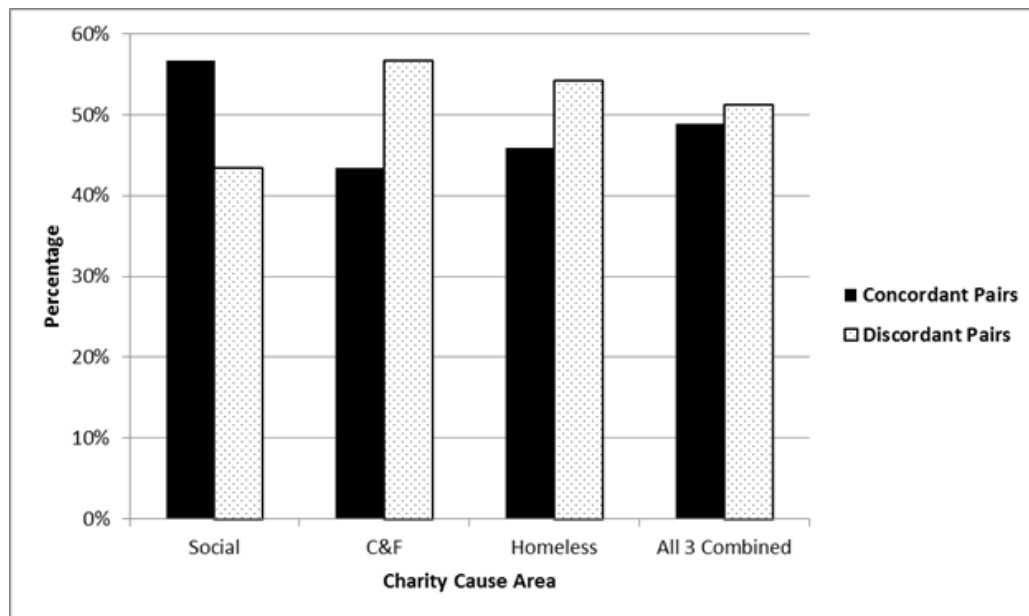
Figure 7.20: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or Two Non-Numerical Scores in Compliance Footprint Question for Element Two -- Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?”



n = 19

Figure 7.21 shows that all three cause areas had high percentages of discordant pairs. Forty-three pairs out of the total of 84 provided different answers (51 percent). For the Homeless Services organizations and Children’s and Family Services organizations, the percentage of pairs who provided different answers is higher than the percentage of pairs who provided identical answers. For Social Services organizations, the percentage of concordant pairs is high (57 percent) relative to other two cause areas.

Figure 7.21: Percentage Breakdown of Inter-Rater Reliability for Element Two -- Test One: “Published Independent Evaluations?” and “Progress Reports on Corrective Actions Undertaken Following the Evaluation?” Researcher Confidence Level for Compliance Footprint Responses



n=84

Test Two: Standards and Certification Mechanisms

This test examines whether a charity holds membership in a standards and certification mechanism and whether that mechanism explicitly covers outcome reporting. These questions were originally written by CN as “Membership in one or more standards and certification mechanisms?” and “Does that standards and certification mechanism explicitly cover outcome measurement and reporting?”

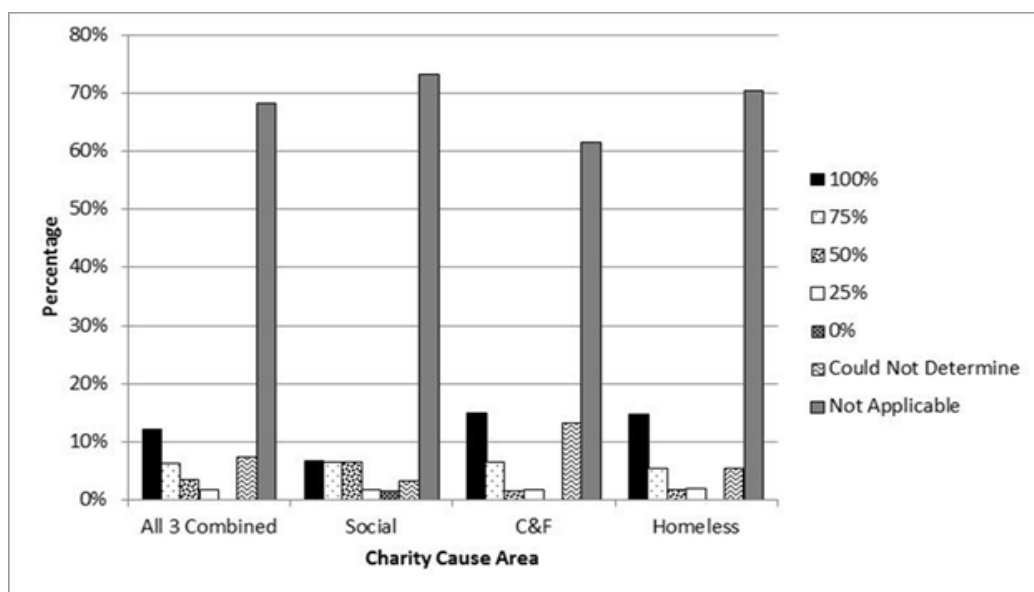
Quantitative: As Figure 7.22 shows, for all three cause areas, students answered ‘Not Applicable’ the most often – 68 percent of the time -- for the test’s Compliance Footprint section. Students selected “Could Not Determine” slightly less than eight percent of the time and “100%,” “75%,” “50%,” “25%,” and “0%” less than 12 percent of the time. As with all three

cause areas combined, “Not Applicable” was chosen the most often for each area independently as well.

Approximately 79 percent (of the 42 responses that show evidence of participation in a standards and certification mechanism) provided enough information to determine Compliance Footprint. These 42 responses, however, only make up about a quarter (24 percent) of the total number of responses (174). In other words, most of charities that are participating in these mechanisms make an attempt to show the degree to which their programs comply with them. But most of the charities surveyed do not report participation in any standards and certification mechanism.

The Capstone Group had a good deal of confidence in their assessments of the Compliance Footprint for the two tests in this Element. The average level for Test One was 2.5 on a scale of 3, and for Test Two, it was 2.4 on a scale of 3.

Figure 7.22: Percentage Breakdown of Element Two -- Test Two: “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?” Compliance Footprint Responses



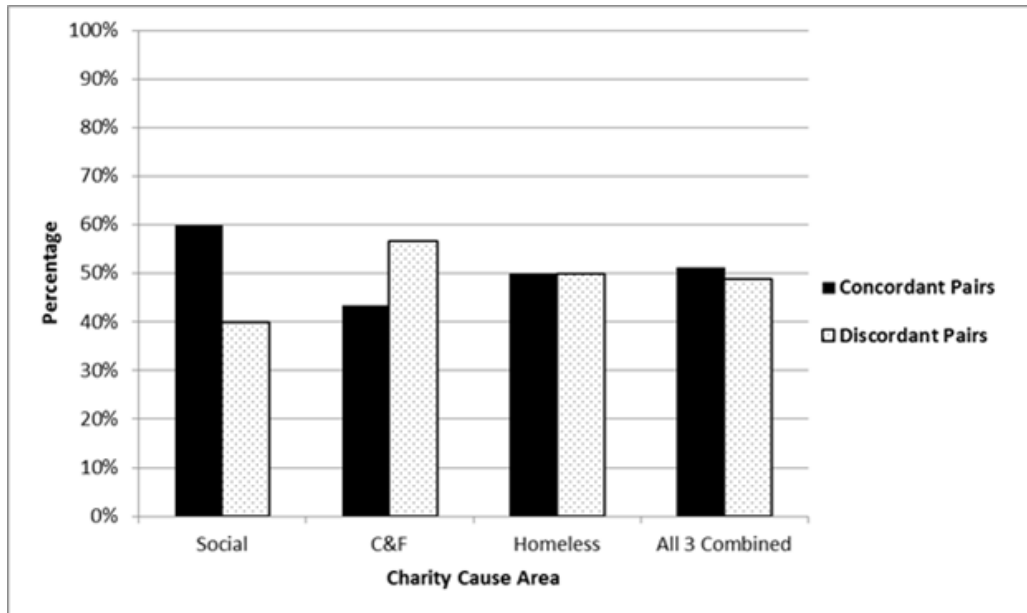
n=174

Qualitative: Students had several concerns about the Compliance Footprint question in Test Two. Primarily, they expressed confusion about what constituted an adequate standards and certification mechanism and an inability to determine what proportion of the charities' programs the mechanisms actually covered. Several students indicated that they did not understand how to apply the Compliance Footprint to this test.

Inter-Rater Reliability: Figure 7.23 shows that for all cause areas, as many student-pairs had different answers for the Compliance Footprint question as had identical answers in Test Two. There are nearly equal numbers of discordant (41) and concordant (43) pairs. For Homeless Services organizations, an equal number of student pairs had identical answers. More student pairs had different answers (57 percent) for Children's and Family Services organizations than had identical answers. Social Services organizations had the highest percentage of concordant pairs (60 percent) for this question, as seen in Figure 7.23.

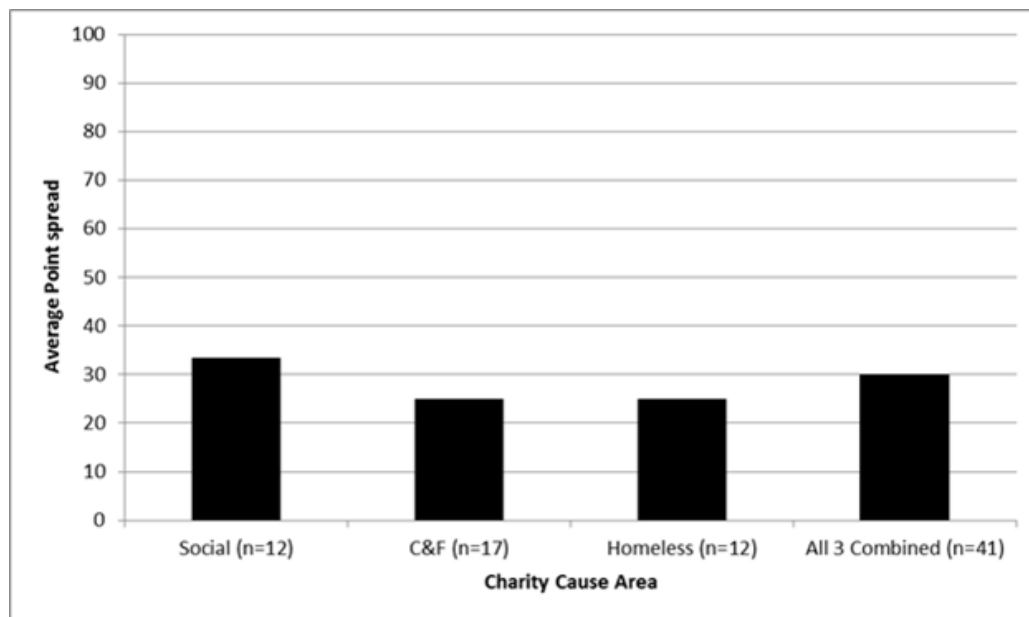
Figure 7.24 shows that average point spread in Compliance Footprint is 30 percentage points for all cause areas. In situations where both students provided a numerical response and there was a difference in how the two students answered, the pair differences were summed and the total difference was divided by the number of discordant pairs. Children's and Family Services and Homeless Services organizations had the same average (25 percentage points), as seen in Figure 7.24.

Figure 7.23: Percentage Breakdown of Inter-Rater Reliability for Element Two – Test Two:
“Membership in One or More Standard and certification Mechanisms?” and “Does that
Standard and certification Explicitly Cover Outcome Measurement and Reporting?” Compliance
Footprint Question Responses



n=84

Figure 7.24: Average Percentage Point Spread in Compliance Footprint Responses of Discordant Pairs for Element Two – Test Two: “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?” (on a scale from 0 to 100)



n = 41

Figure 7.25 shows the number of instances (36 pairs) in which one student in the pair provided a numerical response and the other provided a non-numerical response (“Could Not Determine” or “Not Applicable”) or in which both students provided non-numerical responses, out of the total number of discordant responses (41 pairs) for the Compliance Footprint question. These kinds of discrepancies in responses occurred slightly more often in surveys for Children’s and Family Services organizations (16 times) than they did in surveys for Social Services organizations (9 times) and Homeless Services organizations (11 times).

Figure 7.25: Number of Discordant Pairs with Numerical and Non-Numerical Scoring or Two Non-Numerical Scores in Compliance Footprint Question for Element Two – Test Two: “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?”

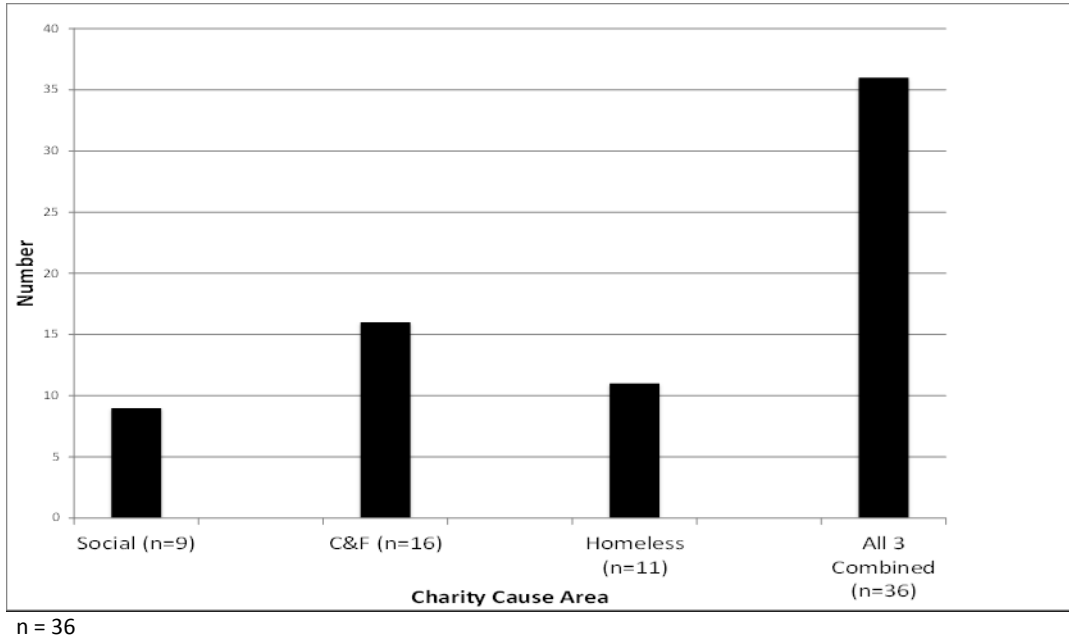
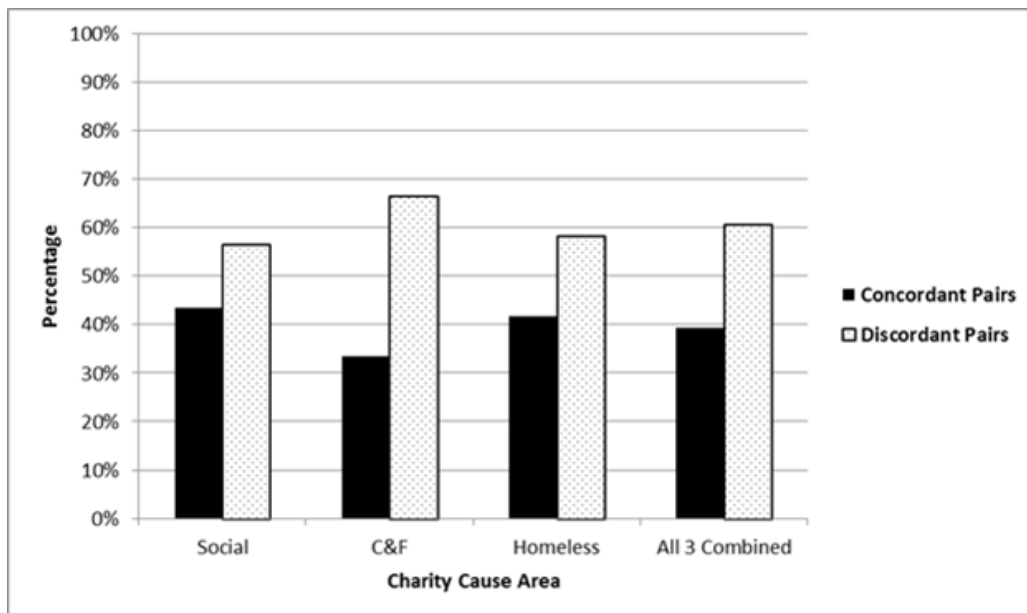


Figure 7.26 shows that both for all three cause areas combined, and each area independently, more student-pairs had different answers about researcher confidence for the Compliance Footprint question in Test Two than had identical answers. Fifty-one pairs out of the 84 provided different answers for this question. Children’s and Family Services organizations had the highest percentage for discordant pairs (67 percent).

Figure 7.26: Percentage Breakdown of Inter-Rater Reliability for Element Two -- Test Two: “Membership in One or More Standard and certification Mechanisms?” and “Does that Standard and certification Explicitly Cover Outcome Measurement and Reporting?”
Researcher Confidence Level for Compliance Footprint Responses



n=84

Implications of Findings for Element Four

The Compliance Footprint question and Inter-Rater Reliability results indicate that students had more difficulty answering the question when CN did not provide an explicit example of how to calculate the Compliance Footprint. The *Charity Navigator 3.0 Research Guidelines* from February 15, 2012, included an example of how to calculate the Compliance Footprint for Test One, but not for Tests Two, Three, and Four, of Element One. As shown earlier, the two most common responses for Test One were “100%” and “75%” program compliance and approximately 63 percent of pairs were Discordant. In contrast, “Not Applicable” and “Could Not Determine” were chosen much more often in the subsequent three tests, and the percentage of discordant pairs increased to approximately 90 percent. Test One,

Is the causal logic plausible? can be inferred from an organization's mission statement, which is often easily found on virtually all charity websites. Information related to dosage and measures to be collected, however, is often more difficult to find on charities' websites, which may have partially accounted for the difference in responses for Test One compared to Tests Two, Three, and Four.

The Compliance Footprint question and Inter-Rater Reliability results for Element Two indicate that students also had difficulty calculating a Compliance Footprint when charities were not forthright about their use of independent evaluations or membership in standards and certification mechanisms. Students often did not know how to apply the Compliance Footprint questions, especially when it was unclear what proportion of a charity's programs standards or certification mechanisms were actually covered, or even where to find such information. The fact that "Not Applicable" was chosen most frequently for both tests and that there were more discordant than concordant pairs for Test One of Element Two may reflect these uncertainties.

These findings suggest that, for Element Four, raters will need to match a charity's activities with program expenses indicated on organizational financial reports, such as IRS Form 990. To ensure that raters do not overlook relevant information, CN's instructions should provide explicit and detailed instructions regarding the website research process. Doing so will not only help raters to determine whether or not charities actually do provide information, such as dosage or measurement collection methods, but also to know which related program expenses to look for on financial reports. As indicated previously, students often had difficulty

choosing between “Yes” and “No” and there was low inter-rater reliability for the binary questions in Tests Two, Three and Four of Element One.

In addition, an issue that arose in all four tests was how to assign a percentage for the Compliance Footprint question when the value fell between the available percentile intervals. In such situations, students did not know whether or not they should round up or down.

Likewise, it may also be necessary to explain carefully how raters can determine whether or not program expenses are specifically connected with a charity’s use of independent evaluations, or participation in compliance standards and certification organizations. Indeed, given the range of activities that many human services agencies conduct, it may be especially difficult to determine meaningfully how much of their expenses are directed towards accredited programs. CN may want to create an explicit framework to guide raters when only a portion of a charity’s programs are eligible for such membership (or capable of being independently evaluated).

The fact that for Element Four, raters will use a pass/fail test, and that charities cannot receive any points for the CN 3.0 results-reporting rating criteria if they fail this test, means that clear instructions about how to deal with uncertainties and find the necessary information will be particularly crucial.

Chapter Eight

Assessment of the Charity Navigator Research Tool

In this chapter, the Capstone Group offers a number of observations about the Charity Navigator Research Tool as a whole and recommendations for improving it. These are based on its experience with the CNRT and the student-developed Background Research Tool, including qualitative data collected in the CNRT's comment box and student journals, as well as discussions among the students throughout the project.

Challenges Encountered in Using CNRT

The Charity Navigator Research Tool relies on results-reporting information that is publicly available on organization web-sites. This is consistent with not only the goal of CN 3.0 – which is to measure how much information about their outcomes charities make available to the public – but also the objective of including a large number of charities in the CN data base. Assessing materials not readily available would be too time- and resource-consuming to be practical, as well as defeat the purpose of encouraging charities to report on their impacts.

However, as the Capstone Group discovered chiefly from the phone surveys, charity websites do not normally provide very much information about certain topics, including internal evaluations or efforts to ascertain constituent views, and sometimes have reasons for wanting to keep such information internal. (Whether these reasons are justifiable or not is a different matter.) As a result, the CNRT may produce incomplete or biased answers. In addition, since the CNRT takes website information at face-value, self-reporting charities can potentially manipulate their ratings by adding – or subtracting – material relevant to the criteria

CN is using. Since follow-up phone surveys are not likely to be feasible once CN 3.0 is operational, other ways of verifying the results-reporting information that charities publish on their websites need to be developed.

It is also important to underscore the limited sample-size for this test of the CNRT. The Capstone Group's research is based on charities from only three cause areas: Social Services, Children's and Family Services, and Homeless Services. (A student group at the Maxwell School of Syracuse University was assessing International charities, but there was no contact with the IU project during the semester.) With no more than 30 charities for each cause area (out of 150 in the CN population for each area), how representative the sample was – and the representativeness of the findings reported here – cannot be ascertained. Students were given no information about how the organizations were selected.

Furthermore, in Element Three, the limitations of the sample size became more problematic because of the use of a screening question: "Does the charity publish rigorously collected feedback data from its primary constituents?" Only 31 responses out of 174 (most organizations were reviewed by two students) continued past this threshold question to the other tests of Element Three. (Only four pairs of responses were in agreement on the answer to this question as well.) As a result, even though other charities may be interested in or responding to constituent voice, the CNRT assessed the practices only of the very small group that were "rigorously" obtaining feedback, who are almost certainly unrepresentative of the charities in the CN data base, let alone the broader population in the three cause areas.

Future tests of CN 3.0 should be sure to examine additional cause areas and use statistically appropriate methods of sampling organizations in each area. Questions that

unnecessarily reduce the number of organizations being assessed should be reframed to ensure that the findings do not conceal outcome-reporting practices that are actually in use.

Since the journal process used by the Capstone Group was unstructured and open-ended, even though it revealed a number of common concerns with the CNRT, it did not record observations for every test for each Element. Indeed, in retrospect, it is clear that certain tests received more attention than others, perhaps because they presented major challenges to large numbers of students. But smaller problems or those noticed only by a few students may have been overlooked in the process. In future tests of the CN 3.0, a more systematic approach to journaling would allow for finer, more nuanced issues to surface. This would enhance the value of the CNRT.

Challenges Encountered in the CNRT Itself

Because of the subjective nature of many of the questions, students often came to different conclusions when reviewing the same charity. Students also had difficulty understanding the intent of many of the tests, i.e., the kinds of information the questions were seeking. Tests that particularly experienced this problem:

- Element One:
 - Test Two
 - Test Three
 - Test Four
- Element Three:
 - Test One
 - Test Two
 - Test Three
 - Test Five

The Capstone Group suggests remedying this by providing a clearer framework and objectives for each element, or each test as needed.

The instructions were also challenging. While the research guidelines document was useful, it would be more beneficial for raters to have clear instructions embedded directly in the CNRT. In addition, a webinar for raters beforehand would help ensure they understood the instructions. The webinar could also provide practical advice on the most likely places to find information on charity web-sites, how far afield (or in-depth) volunteers should go to find the necessary information, and how recent the information should be in order to meet each Element's tests. (For example, it was not clear whether students were required to review the entire website and if they should review auxiliary documents such as Annual Reports – and for which years? -- or Newsletters for results reporting information.) Tests that particularly exhibited this problem:

- Element One:
 - Test One
 - Test Two
 - Test Three
- Element Two:
 - Test One
 - Test Two
- Element Three:
 - Test Two
 - Test Five

Many students cited language and terminology as one of the major problems with the tool. One concern was that the wording was not always easily understood by a diverse group of individuals. (Members of the Capstone Group included students with little knowledge of

nonprofit organizations and international students.) Many also felt that the questions contained nonprofit jargon that would be difficult for more typical raters than a class of MPA students to understand. Some felt that terms were not clearly defined (and could have multiple meanings).

To address this, the research guidelines and the CNRT should be written as simply as possible, with minimal use of technical terms, and as necessary, provide definitions that leave little room for interpretation. The CNRT might also include a pop-up dictionary that a rater can use for words that are difficult to understand, or to provide a hyperlink to Charity Navigator's or Keystone Accountability's glossary of terms. Tests that particularly experienced this problem:

- Element One:
 - Test One
 - Test Two
 - Test Three
- Element Two:
 - Test One
 - Test Two
- Element Three:
 - Test One
 - Test Two
 - Test Three
 - Test Five

Many students would also have welcomed including examples in the research guidelines to illustrate terms or the kinds of responses that would satisfy the tests. For instance, for some of the binary questions, Charity Navigator could provide examples of what kinds of reporting should be given “Yes” (or “No”) answers. On the questions asking for comparisons, examples of

organizations that might be appropriately compared would have been useful. These illustrations could be placed either in the instructions or, via pop-ups, in the CNRT itself. Tests that particularly exhibited this problem:

- Element One:
 - Test One
 - Test Three
- Element Two:
 - Test One
 - Test Two
- Element Three:
 - Test Four
 - Test Six

Many of the tests were structured so that students felt their answers were constrained or limited and did not allow for a response that captured the intent of the question. Requiring binary responses for tests that deserved a scale or open-ended answer was a frequently noted problem. While binary questions lead to easier coding and interpretation, many students felt that a “Yes” or “No” did not always give an accurate portrait of what the organization being assessed was doing. For example, asking whether or not an organization “rigorously collected” feedback data did not allow students to indicate that there was a system for collecting feedback, but it was only “somewhat rigorous.” A “somewhat rigorous” system would result in either a “no” response because it was not a “very rigorous” system, or a “yes” response because the system made an attempt to collect at least “somewhat rigorous” feedback. A rating scale, permitting a score for the degree of rigor, would be a more appropriate structure for this, and, *mutatis mutandis*, other tests.

Another structural problem the Capstone Group encountered involved the consistency of questions among the four Elements. For example, Element One and Element Three have only one Confidence Level question for each test, while Element Two has multiple Confidence Level questions for one test, while the other test in Element Two has a Confidence Level question only on the Compliance Footprint. Also, certain Elements contain questions with unique formats, such as the threshold question in Element Three and the binary tests in Element Two that are structured as statements. Likewise, Elements One and Two both contain Compliance Footprint measures, but Element Three does not. While the Elements are measuring unique concepts, the lack of consistency in the questions the CNRT poses could create confusion among the raters and produce uneven findings about the charities. Tests that particularly experienced this problem:

- Element One:
 - Test Two
- Element Two:
 - Test One
 - Test Two
- Element Three:
 - Test One
 - Test Four

The Particular Challenges of Element Three

The preceding comments apply to all the Elements. But Element Three: Constituent Voice is the most novel feature of CN 3.0. Not surprisingly, it presented some challenges of its own, including how best to conceptualize it.

As noted above, the first question in Element Three is unique because it is a screening or threshold question. The Capstone Group encountered problems in its wording, format, and order.

Many students mentioned in their journals that determining when feedback data were “rigorously collected” could be very subjective. To address that, CN should try to specify the meaning of “rigorously collected,” perhaps by concrete examples or detailed checklists of the kinds of feedback data that fit the criteria within the question itself. Likewise, identifying the “primary constituents” of a charity is not always a straightforward matter. CN should provide parenthetical definitions in the test, particularly in ambiguous cases, such as “Children’s and Family Services” agencies, which may have at least two sets of “primary constituents.”

As indicated previously, the binary and threshold nature of the first question proved problematic. A “Yes-No” answer does not allow raters to provide a complete picture of how rigorously feedback data are collected and should be replaced with a scale that would allow raters to report the degree of rigor. This would also allow charities whose feedback collection methods are less rigorous to still be assessed on the other tests of Constituent Voice.

That could be particularly important for furthering the developmental goal of CN 3.0. A charity transitioning into including constituent voice among its outcome measurements would thus have an incentive to try to pass the other tests in Element Three, even if its system to collect feedback data was not yet “rigorous.”

Revising the wording of Test One as follows, and changing its status as a screening question, would achieve this objective:

- Does the organization describe, reference, or show the results of collecting aggregate feedback from its primary constituents (direct recipients of services)?
- How rigorous is the system of collecting feedback?

The second test for Constituent Voice – “Is that published feedback data presented in a way that shows changes over time going back at least one year? -- presented similar problems. The wording and instructions confused many students, according to the journals and class discussions. For example, they were unsure whether the test was looking for charities to have conducted a pre- and post-intervention evaluation, or for charities to have conducted an annual evaluation not tied to the progress of specific clients. CN needs to explain and refine the question to better fit its intention.

Furthermore, the Capstone Group was concerned that the time-frame specified -- “going back at least one year” —was arbitrary. Some charities might engage in “regular” or “consistent” feedback collection, but on a quarterly, semi-annual, or annual basis. Some charities also expressed concerns about the expense and difficulty of collecting long-term data for programs that have high client turnover.

By revising it as follows, the test can overcome these problems:

- Does the charity regularly (quarterly, semi-annually, or annually) engage in feedback collection from primary constituents?
- Does the feedback data track differences in constituent feedback over time?

What “represents” means loomed large in the third test, which seeks to find out how well reported feedback reflects the nature of the people a charity serves. Students said that the question directed the rater to think about the *quality* of the data provided--how accurately

it represents the constituents' feedback--instead of whether the charity provides information about sample composition and the like that could put the feedback in context. If the intention is simply to provide CN users with evidence about sample characteristics, the question could be revised as follows:

- Is there a description or statement about how representative the published constituent feedback data is in terms of gender, ethnicity, age, etc.?

The Capstone Group did not find many problems with the fourth test (apart from the fact that charities which did not "rigorously" collect feedback data could still derive information about program effectiveness from what they did obtain, but would not have been judged on this question). Although some students were confused as to whether "effectiveness" referred only to positive contributions, or might also include negative results, the principle difficulty they had with this question was knowing how thoroughly to look for answers. CN might provide examples of where the answers would likely be found on a charity's website and the kinds of answers raters should be looking for.

As with other questions, students had concerns about the wording and intent of the fifth test, which concerns transmitting feedback information to constituents. Based on the journals and discussion groups, many students recognized that "reporting back" could take many forms and wanted to better understand which should be judged acceptable. (E.g., is posting feedback results to the website sufficient? Printing them in an annual report?) In addition, self-servingly or not, charities advised the Capstone Group that some feedback information could not be publicly reported because of privacy or other legitimate concerns. A small revision, as follows, could make this question more useful:

- How does the organization share results of feedback (through community meetings, flyers, or other indicators) with its primary constituents, whom the information affects?

The last test, student felt, was really asking for information about two topics: whether or not a charity was collecting feedback similar to that collected by other charities and if so, whether or not it was comparing its data with that of the other organizations. Conceivably, charities could do one without doing the other. Furthermore, the CNRT gives no guidance about what organizations might be considered “comparable” for a particular charity. CN needs to clarify in the instructions what an acceptable comparison is, including whether other chapters of a larger organization count. It could also include an open-ended question to allow raters to list the comparison charities to ensure that they are relevant to the one being rated.

Following is a possible way of revising this test:

- Does the organization standardize its questions with similarly-focused organizations to create comparable feedback?
- Please list organizations used as benchmarks for comparing feedback data.

Additional Tests

In addition to analyzing each Constituent Voice question for its strengths, weaknesses, and validity, the Capstone Group also looked at them in light of the theoretical concept of constituent voice. It relied on Keystone Accountability’s website and white paper, *21st Century* (Bonbright, Campbell, & Nguyen, 2009) to develop a framework of concepts that are essential for a charity to demonstrate its effective use of constituent feedback. The current CNRT

questions were compared to this framework to see what gaps might exist and additional test questions were then created to fill those gaps.

According to *21st Century*, effective use of constituent voice is built around four components: Value, Systems, Analysis, and Utilization.

By **value**, the paper means that a charity should be able to demonstrate that it values both the needs and opinions of its beneficiaries regarding its ability to serve them, which includes increasing their capacity for participating in opinion-based empowerment activities, like providing feedback (Bonbright, Campbell, & Nguyen, 2009, p. 16). Many charities state that they value constituent feedback, but, they should also be able to demonstrate it by enhancing the ability of their constituents to provide it. (Bonbright, Campbell, & Nguyen, 2009, p. 22).

By **systems**, *21st Century* is referring to a charity's having reliable, systemic policies or mechanisms for collecting feedback, which, the paper says, are becoming more common in practice (Bonbright, Campbell, & Nguyen, 2009, p. 21).

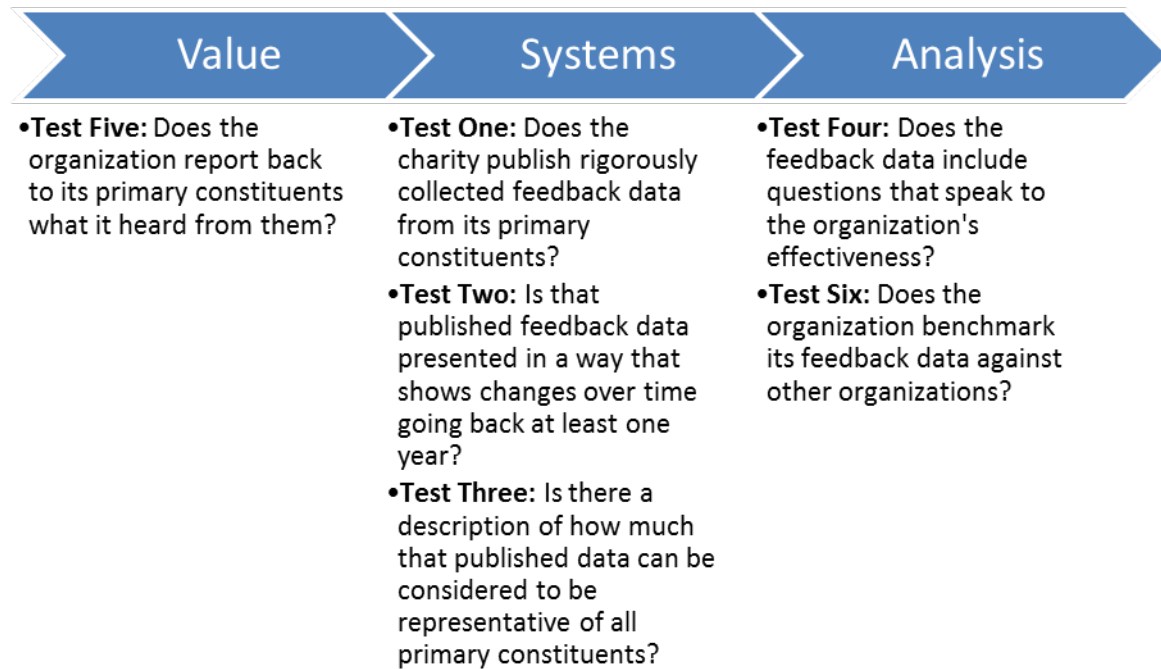
By **analysis**, *21st Century* means that a charity must be able to examine the feedback in ways that reveal meaningful trends, ideas, and needs to the charity, which all start with useful measurements. The paper found that, overall, most charities do not have consistent or meaningful measurements that enable them to achieve helpful analysis and that rectifying that is vital for charities to move forward with constituent feedback (Bonbright, Campbell, & Nguyen, 2009, p. 23).

Finally, by **utilization**, *21st Century* refers to how a charity demonstrates that it incorporates feedback components into programs for improved or more effective services. This

component addresses the uses of constituent feedback identified in the paper, specifically for “improvement” and “strategy development” (Bonbright, Campbell, & Nguyen, 2009, p. 16). It could also loosely include “civic engagement” and “societal learning,” although these concepts may be more appropriately included in the “value” section (Bonbright, Campbell, & Nguyen, 2009, p. 16).

Against this framework, the Capstone Group matched the test questions to determine how much of what Keystone Accountability proposes the CNRT is addressing. Figure 9.1 shows the result of this exercise:

Figure 9.1: Test Groupings for Element 3



It appears that Keystone Accountability's components of "systems" and "analysis" are fully satisfied by the current test questions, but that "value" may be under-addressed and "utilization" is not addressed at all. While reporting feedback results back to primary constituents would show that the charity is analyzing the feedback, utilization refers more specifically to implementing policy or programmatic changes based upon the feedback data. Given that the biggest hurdle to effective use of constituent voice now, as reported in *21st Century*, is the establishment of comparable and useful measurements, it is understandable why there is so much emphasis on the components that would most readily do so.

However, under-addressing the other essential aspects could be problematic for two reasons. First, it runs the risk of not capturing current demonstrations of feedback practices in charities, which, as discovered in the phone interviews, already poses a problem to the validity

of using CNRT to measure the concept of constituent voice. Second, only focusing on the measurement and analytic components of constituent voice may end up influencing these specific practices only, discouraging using all of the promising benefits that can be derived from the collection and utilization of constituent data.

Three additional Constituent Voice test questions could be added to better address and encourage the components of “value” and “utilization.”

The first (or seventh test overall) would be:

- Does the organization have a staff or department directory listed with contact information for more than one staff or board member?

The intent of this test is to assess how much charities value primary constituent feedback, since charities that do so should be more forthcoming about contact information for their clients, donors, and other constituencies. Transparency and accessibility are among the many ways for charities to become approachable for their primary constituents. When conducting background research, the Capstone Group found that certain charities provide only an online form to submit comments or contact the charity. The stronger the charity’s commitment to constituent voice, the more it will adopt accessible and appropriate systems for “feedback providers to participate and communicate their opinions,” recommends *21st Century* (Bonbright, Campbell, & Nguyen, 2009, p. 16). Using a measure of capacity to receive feedback shows how a charity enables constituents to contact it. However, passing this test should require more than just a “contact us” section on its website that lists a secretary and PR person’s email, but no other staff members’ contact information. It should have contact information for multiple people, including, possibly, board members.

The second (or eighth) test proposed is as follows:

- Does the organization state a core commitment to downward accountability, being member-led, and other related concepts?

The intent of this test is similar to the previous one in measuring how much charities value primary constituent feedback. But while the previous one offers an implicit indication of the value of constituent voice by asking about accessibility, this one asks for a written, explicit commitment to utilizing it. Organizational statements endorsing downward accountability, being member-led, and themes of empowerment displayed in the mission, values, and programs offered by charities, would indicate a high value placed on bottom-up approach to feedback, consistent with the idea of constituent voice. However, the Capstone Group's research showed that only 34 percent of charity organizations listed an explicit policy or value relating to feedback. Although terms like "downward accountability" and "member-led" need to be defined for the raters in the instructions, a question calling for charities to demonstrate a commitments to them voice might encourage improvement.

The last (and ninth) test aims at the utilization framework as well as the strategy development purpose of feedback, where "feedback data can be used to adjust program strategy" (Bonbright, Campbell, & Nguyen, 2009, p. 16). It might be worded as follow:

- Is there evidence that the organization has used or uses primary constituent feedback to make program or policy change(s) in its service delivery?

The goal behind collecting constituent feedback data is to utilize the information in the practical implementation of changes in program services. Yet the CNRT does not make any attempt at measuring if a charity does that. While many charities may not report their use of primary

constituent feedback in evolving program or service delivery, those who do should be recognized for doing so, since that is the ultimate goal behind the concept. The problem is how much evidence does a charity need to show to pass the test. Is even a minimal policy change enough? In addition, what constitutes “evidence” would need to be defined in the instructions. Nonetheless, without an effort to get at the utilization component, CN 3.0’s effort to assess Constituent Voice among charities in its data base would be incomplete.

Chapter Nine

Utility of the CN 3.0 Ratings Tool: Findings from a Survey of Donors

A critical assumption of CN 3.0 is that donors will want and use information captured by each element when making a decision to donate to a charity. To determine the overall importance of the CN Elements to a potential donor's decision to give to a charity, the Capstone Group conducted a short survey of donors to CN who gave 25 dollars or less in the last year. The results produced two primary findings: first, few donors currently used the kind of information to be included in CN 3.0 to make giving decisions, but many were interested in information regarding each element; and second, donors were clear that they use CN as a trusted source of information, often in place of the charities themselves. Two interesting themes arose throughout the free response comments: first, donors expressed concerns about and limitations regarding the measurement and publishing of information on the CN Elements (which were also noted by charities themselves in the phone interviews and by CN in their concept note); and second, donors were worried about charities' publishing inflated results to earn a higher rating.

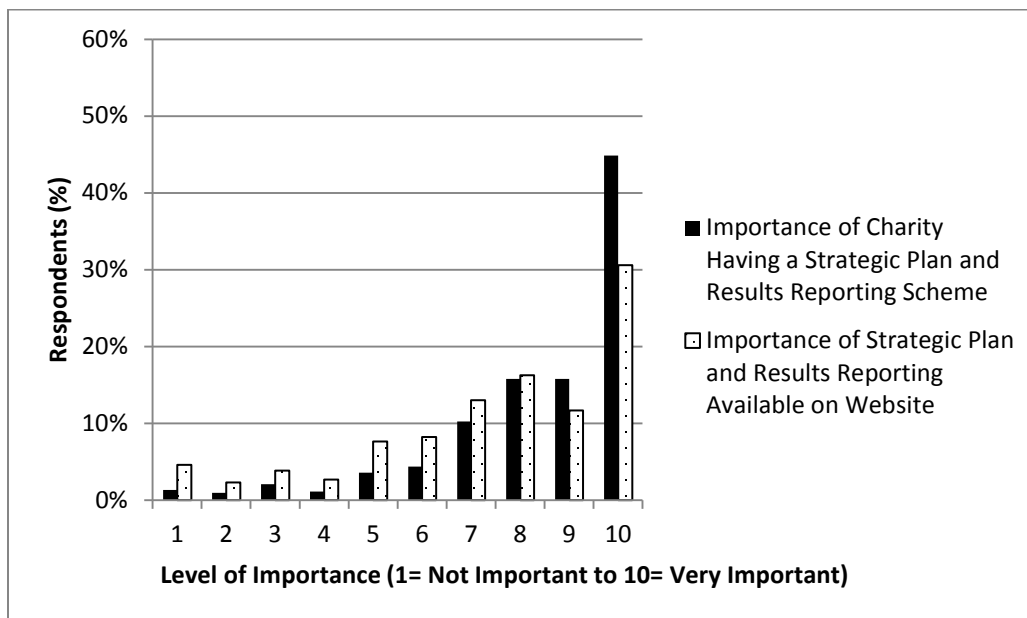
How Important is Strategic Planning?

Question One of the survey inquired as to whether it was important to the donor that a charity has a strategic plan and a plan for program evaluation when donors were making their decision to donate. Part A of the question specifically asked how important it was when making a decision to donate that, "the charity had a plan that described what programs it provides and what they measure to be certain the programs get good results." Part B of the question asked

how important it is to potential donors that “the charity displayed the plan on its website.”

Figure 10.1 below summarizes the responses to this question.

Figure 10.1: Importance of Having a Strategic Plan and Plan for Reporting Results



n=526 for having plan
n=523 for displaying plan

Responses for Parts A and B of Question One cluster around the upper extreme of the response scale, indicating a good portion of donors care that an organization has a strategic plan and a plan for program evaluation. Nineteen of 83 respondents writing in the free response section in Part A emphasized that these pieces of information were important to their decision to donate. Somewhat fewer respondents care that this is published on the charity’s website, but they did desire access to this information. Ten of 51 mentioned this point in the free response section.

An analysis of the comment section for Question One Part A shows that donors rely, to a great extent, on CN to assess details of charities’ Logic, Results, and Measures, as well as

Independent Evaluations and Standards and Certification Measures. For example, one respondent commented: “I depend on Charity Navigator to obtain and provide such information [not the charity itself]” [authors’ clarification]. Six of 83 free response comments echoed this, citing that they do not check charities’ websites or that they defer to CN on such matters.

As evidenced by clustering of responses in the upper response range (approximately 91 percent of total responses), CN donors value information on all the 3.0 Elements. One point to emphasize is that though donors are, on the whole, concerned with whether the charities they support produce results, they also understand how difficult it is to measure, evaluate, and publish that data. One respondent summed it up nicely: “I know how hard it is to measure this kind of result. And how much people spend on such efforts. May not be worth the money [for the charity].” CN’s communication channels and educational resources, outlined in the Concept Note, will be critical to addressing these concerns for both donors and charities.

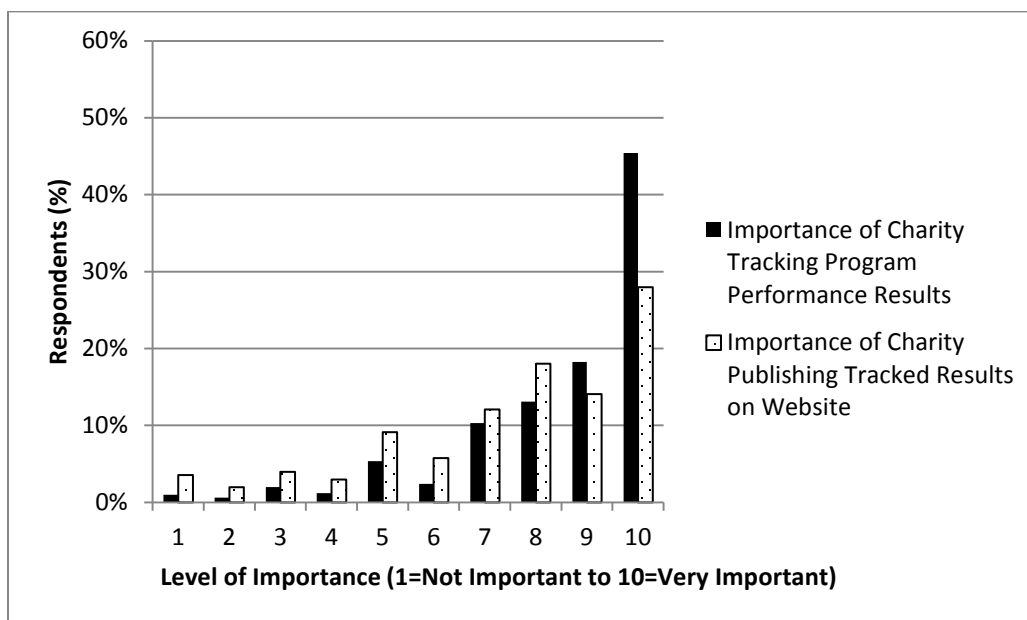
Part B inquired as to whether it was important in making a decision to donate that a charity display information about results measures and results reporting on its website. Though the majority of respondents wanted access to this information via the charity’s website, (80% of respondents for Part B in the upper range) it seems to be somewhat less important than results measures in the first place.

How Important is Tracking Results?

Question Two of the survey asked whether it was important that a charity track the results of its programs when making a decision to donate. Part A of the question specifically asked how important it was that, “The charity kept track of the results of its programs.” Part B

of the question asked how important to potential donors it was that “the charity displayed how it tracks results on its website.” Figure 10.2 below summarizes the responses to this question.

Figure 10.2: Tracking Performance Results



n=504 for tracking results
n=504 for displaying results

Responses for both parts of Question Two cluster around the upper extreme, with 89 percent falling in between 6 and 10 on a 10-point response scale, which seems to indicate a large proportion of donors care that a charity tracks its results. With regard to whether the charity displayed how results were tracked on its website, respondents seem to feel positively about a charity doing so (78 percent chose 6-10 on the response scale).

Analysis of the free response comments indicates similar attitudes as revealed in Question One, Parts A and B. In particular, donors indicated that they relied on CN to track and provide this information for them, as one respondent put it, “I don't usually check [charities’] websites” (this answer occurred in 14 of 41 free responses to Question Two). More than for Question One, respondents answering Question Two felt that tracking program results was an

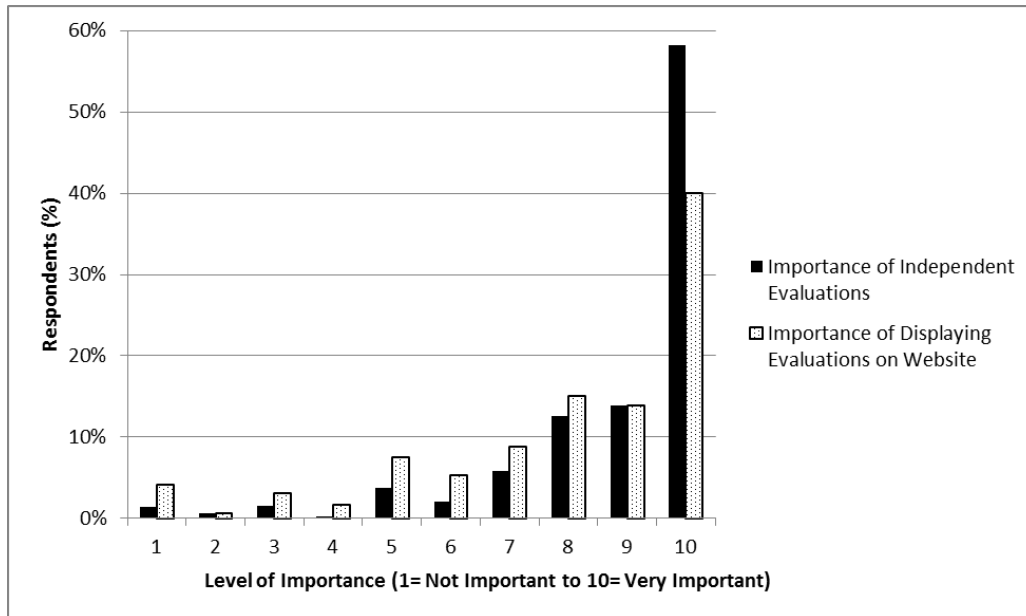
obvious component of their decision to donate; as one respondent put it, “This is just too obvious to comment on, right?” (10 of 44 free response comments for Question Two). Other respondents noted challenges to measurement for charities, the need for indicator development by CN, and the possibility of falsifying information (“cooking the books”).

How important are Independent Evaluations?

Question Three of the survey asked for the donor’s perspective on the importance of externally conducted independent evaluations when making funding decisions. Part A asked donors, “When you've given to charities in the past, how important was it that the charity's results had been evaluated by someone who is not affiliated with the organizations?” Part B asked how important it was to donors that the charity shared the results of the independent evaluation on its website.

Nearly 60 percent of the 498 respondents rated the existence of these independent evaluations as “Very Important.” As the Figure 10.3 below demonstrates, the answers to both questions were clustered toward the upper half of the scale, with a noticeable difference in the percentage of respondents that found each question “Very Important.” It appears that while most respondents found the existence of the evaluations to be critical, it was less important that this data was placed on the website.

Figure 10.3: Importance of Collecting Displaying Independent Evaluations



n= 499 for conducting evaluations

n= 490 for displaying evaluations

The free response comments reveal these distinctions as well. While 17 of the 60 comments stressed the importance of evaluations, a few respondents commented that the independence of the evaluator does not necessarily imply greater competence. As one respondent put it, “Although an outsider is generally preferable, I have known a few organizations to build good, honest in-house capabilities, although usually with some outside assistance.” This respondent also stated she felt the critical issue was the ability of the evaluator. It seems reasonable to conclude that donors find considerable value in evaluations and see evaluations as an important tool for increasing their trust in an organization. While a few do not necessarily believe that the independence of the evaluator is the most important

factor, donors seem to appreciate efforts to conduct objective evaluations and publish those results.

Still, many respondents were sensitive to the needs and capacity of charities. Twelve of 60 comments indicated some sensitivity toward the perceived resource and financial costs of these evaluations. Respondents expressed concern that evaluations might not be feasible for all organizations, notably smaller charities.

For the second part of the question, whether charities should post evaluation results on their websites, 10 out of 47 free response comments indicated that donors are not searching charity websites for this type of information. Many commented that they considered it the job of outside organizations like CN to provide information about the results of evaluations, not the charities themselves. Nine out of 47 comments mentioned that they would rather see evaluations on an external site like CN or through other means than the charity's website.

A number of comments that expressed skepticism regarding the quality of evaluation results posted by charities support these results. As one comment stated, "I don't want organizations to pick a simple or glitzy stat just to put it on the website." The donors in this survey seemed to be concerned about falsified results, and the independence of the evaluations does not always allay these concerns. It appears some donors examine even objective, independent evaluations with a critical eye. In short, it seems donors do find value in independent evaluations, though they have some concerns about the objectivity of the evaluations and whether organizations of all sizes can afford them.

How Important is Constituent Feedback?

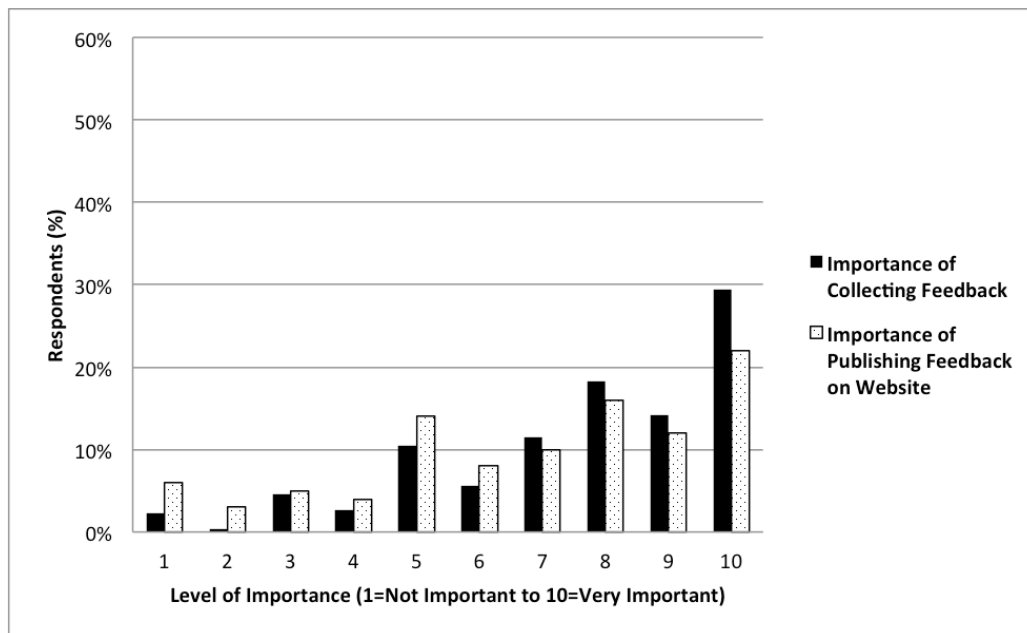
The fourth set of questions in the survey addressed the value to donors for their giving decisions of charities' collecting and displaying feedback from primary constituents. Part A of Question Four asked the donors how important it was that, "To help it improve its programs and services, the charity collected and used feedback from the people or groups it serves." Part B of the question asked how important it was that, "The charity displayed the feedback and responses on its website."

As displayed in Figure 10.4 below, 486 and 479 people responded to the two parts respectively. In general, responses indicated that constituent feedback is important to donors when making decisions. Twenty-nine percent of respondents indicated it was "Very Important" to them that a charity collected such information. Most felt that it was in some way important, with levels six through ten comprising 79 percent of the responses, and levels one through four comprising just ten percent of the responses. Respondents did not land completely at the extreme, though; levels seven through nine each received significant portions of the responses, indicating that while this concept may be important to many donors, it may not be one of the most important.

The responses to Part B were also skewed toward "Very Important," but at lower levels than the importance indicated in Part A. Twenty-two percent of respondents indicated "Very Important" for this question. Levels six through ten comprised 68 percent of responses, while levels one through four comprised 18 percent. This flatter distribution indicates that while

donors do consider it important for charities to publish feedback, it is not as important as actually collecting it.

Figure 10.4: Importance of Collecting and Displaying Constituent Feedback



n=486 for collecting feedback

n=479 for displaying feedback

Overall, respondents indicated that the collection of constituent feedback is an important practice for charities and that the publication of this feedback is also meaningful to donors, but to a slightly lesser extent than other activities, such as strategic planning and tracking results. It appears that CN has the support of its donors in its decision to evaluate the presence of constituent feedback practices.

However, this support may be tempered slightly by a few considerations. First, the structure of the question implied a very positive outcome (“to help it improve its programs and services”), which may have led respondents to focus on the outcome rather than the process of collecting feedback itself. As outlined in *The 21st Century Potential of Constituency Voice*

(Bonbright et al., 2009), improving programs and services are only two of the reasons a charity might collect feedback from its constituents. Other beneficial, but less obvious ones might include accountability to funders, building capacity among beneficiaries, civic engagement, and societal learning. If respondents were favorably inclined toward collecting constituent feedback because they saw it chiefly as means of program improvement, the high levels of support shown in the survey might be misleading.

The second consideration complicating the survey's findings is the nature of the free response comments, which indicated some misinterpretation of the questions. Fourteen of 44 respondents to Part A commented that they had either never thought of this concept or had never seen it in practice, making it very difficult for them to imagine it impacting their giving decisions. These types of comments, which might also reflect the views of other respondents who did not comment, suggest that some respondents were not expressing a considered opinion about the value of constituent feedback.

Free-response comments also expressed concerns about the practicality of obtaining feedback or its publication. For part A, six of the 44 respondents indicated that this concept is impractical for many charities to operationalize, with another two saying it was not applicable to some organizations, such as nature-oriented charities. Additionally, some respondents voiced concerns about the possibility that a "vocal minority" could unduly impact organization decisions, or even questioned the legitimacy of the practice: "The fact that people are receiving a free benefit doesn't entitle them to critique it. That would be like biting that hand that feeds you." This question of legitimacy extends to concerns that feedback would inevitably be more political than objective (3 out of 44).

For Part B, six out of 39 commented that it did not matter if feedback was present on the website, but only that it was published somewhere or would be available upon request. Four out of 39 respondents said that they did not trust individual charities to publish unbiased results on their website. Another four out of 39 indicated that they relied on Charity Navigator for their decisions, not the charity's website. Three out of 39 indicated never visiting the charity's website.

The rate of favorable responses greatly overshadows the critical comments outlined above. In moving forward, CN can confidently assume this Element to be important to its donors and their decisions to donate. However, it may also wish to address some of its donors concerns through public education efforts. In comparison to the other Elements addressed in the survey, constituent feedback responses did not distribute as clearly toward "Very Important." This difference, in combination with the many comments indicating unfamiliarity with the concept, suggests that there are significant opportunities for increased education on the need, process, and benefits of collecting and utilizing constituent feedback.

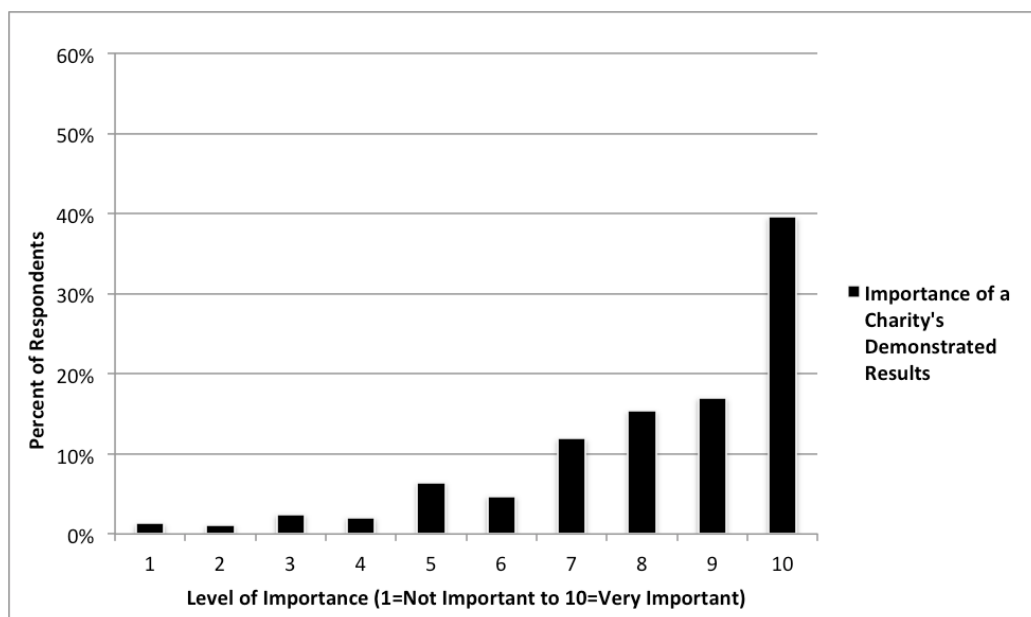
How Important is Results Reporting?

The fifth question was intended to be a summary question that assessed the overall degree of importance donors placed on results reporting when making funding decisions. The question asked if, "It mattered to me that an organization was able to demonstrate results before I made a funding decision."

The survey found that most donors consider demonstrated results to be important, with the results skewed toward "Very Important." As Figure 10.5 indicates, 89 percent of responses identified the importance of demonstrating results as six or higher. Additionally, 38 percent of

responses ranked it as being “Very Important,” the highest level of importance, and another 17 percent ranked it as a nine out of ten overall in level of importance.

Figure 10.5: Importance of Demonstrated Results



n=484 for demonstrating results

Free responses from those surveyed indicate that donors are generally concerned with results reporting, but other factors may be important in their decision, in some cases, superseding interest in results. Of the 78 free response comments, the most common concern about charities related to their financial transparency. Sixteen respondents were interested in information that would help them determine whether or not charities were using an acceptable percentage of revenues to fund programs or provide services, and not paying princely sums to executives. Also, sixteen respondents worried that results reporting could “crowd out” newer organizations and expressed commitment to helping new organizations with promising ideas and/or management teams get started. Additionally, ten respondents commented that it would be difficult for many charities to demonstrate results because of their missions and the

clients they served. They expressed the view that an organization's purpose or mission might not be conducive to showing results within a few years, as might be the case with a cancer research organization. Thus, results orientation and expectations needed to be flexible in order to accommodate different organizations and purposes.

Caveats, Conclusions, and Recommendations

The preceding analysis suggests a number of caveats, conclusions, and recommendations for CN 3.0:

Caveats:

- Due to the characteristics of the sample used for this survey, results should not be generalized to any other donor populations.
- Limitations in survey design and implementation (discussed in the Methodology section above) may have affected the quality of the survey results. These should be considered when interpreting the data.

Conclusions:

- All questions show clustering around the upper end (between response options 6 and 10) of the response scale. Part A of each question tended to have more clustering around response options six to ten than did Parts B. This indicates that donors can see themselves using information on the 3.0 Elements to make donation decisions, but do not see themselves gathering it from charity websites.
- Overall, donors were most interested in seeing charities use and report results measures (91% selecting six through ten). Free response comments showed that donors were aware of the myriad challenges in collecting, measuring, and evaluating such information that could be barriers for many charities. A somewhat smaller proportion (80% selecting six to ten) of donors wanted results measures to be reported on the websites of charities.

- Donors expressed the view that they find value in a charity's use of independent evaluations. Yet, a number also expressed hesitation about the potential costs of such studies, especially for smaller organizations. Others were careful to note that *objective* evaluations could suffice in the place of third-party evaluations.
- Donors acknowledged the potential value of constituent voice in improving how a charity operates. But, they also had concerns over its use in making decisions to donate to charity. Specifically, they could not understand how it would be measured by many charities or could inform their decision to donate.
- Nearly nine out of ten donors value organizations' demonstrated results when making a decision to donate. Yet, free response comments suggest that donors have some reservations about how much emphasis to place on results. They expressed the view that other factors, like financial transparency and an organization's vision, mission, or purpose, might be of greater importance to them in some cases.
- For all five questions, respondents were concerned that charities might falsify or glamorize measures.
- In general, respondents indicated that they depend on CN to provide a judgment on performance, evaluation, and feedback data, suggesting a consumer base for CN 3.0 information.

Recommendations:

- Set a tone of collaboration -- Create or maintain transparent communication channels between CN, charities, and potential donors to address concerns and provide capacity-building resources.
- Create and distribute educational resources emphasizing the importance and utility of the CN 3.0 Elements and methodology. These should be tailored to both donors and charities.

Chapter Ten

Case Study: AmeriCorps Improving Health Throughout Indiana

To further test the utility of the CN 3.0 framework, the SPEA Capstone Group sought to apply it to assessing seventeen organizations participating in the AmeriCorps Improving Health Throughout Indiana (AIHTI) program. AIHTI is an “AmeriCorps State program sponsored by the Department of Applied Health Science at Indiana University Bloomington” (AIHTI 2011-2012 Grant). It serves as a donor to the seventeen host sites, providing manpower (AmeriCorps members) to run health-related programs at each site.

AIHTI began in 2007 and has since paired over 150 AmeriCorps members with over 40 host sites in South-Central Indiana. Current AIHTI members range in age from 17 to 71 years old. Members fill positions which require 300, 900, or 1700 hours of service annually.

AIHTI itself receives grants from the Federal Government for its programs and must apply for these grants each year, as well as evaluate their results. The 2011-2012 grant identifies the compelling community need for AIHTI by explaining that “obesity and being overweight are major causes of serious diseases such as diabetes, heart disease, stroke, cancer, arthritis, and dementia” (AIHTI 2011-2012 Grant). In the state of Indiana, which is the fourth most obese state in the U.S., obesity rates have increased from 46 percent in 1990 to 62 percent in 2004, with over 30 percent of young people already being overweight (AIHTI 2011-2012 Host Site Fact Sheet).

The overall goal of AIHTI is to “help eliminate health disparities in Indiana by promoting healthy weight management through proper nutrition and physical activity” (AIHTI 2011-2012 Host Site Fact Sheet). Each host site has a distinct program, with its own specific objectives, for

an AmeriCorps member to manage. Programs range from weight management to nutrition education and impact groups from youth to seniors.

AIHTI's 2011-2012 placements were at the following 17 sites:

- Indiana University
- Area 10 Agency on Aging
- Boys & Girls Club of Bloomington
- City of Bloomington Parks and Recreation Department
- Deaconess Family Medicine Residency
- Ellettsville Boys & Girls Club
- Girls Incorporated of Monroe County
- IU Health
- Jameson Camp
- Monroe County Health Department
- Monroe County YMCA
- New Hope Family Shelter
- Rehabilitation Hospital of Indiana
- Ruth Lilly Health Education Center
- Stepping Stones
- Volunteers in Medicine
- Youth Services Bureau of Monroe County

Applying CN 3.0 to AIHTI

To improve its organizational capacity, every three years, AIHTI is required to have an outside individual or entity review some element of its program. This year, AIHTI chose to review the evaluation process because it realized that it was a weak spot for the organization. In August 2011, it contacted SPEA to request to serve as a spring semester capstone client.

This request presented an opportunity to test the utility of the CN 3.0 Elements in a public-private partnership. Charity Navigator sees its framework as potentially increasing the program information available to donors and incentivizing charities to focus more on results-reporting. Government programs (such as AmeriCorps) also face pressures to be more outcome-oriented in their funding decisions, just as their grantees are increasingly expected to measure what they are accomplishing. If the CN 3.0 Elements could be successfully applied to help AIHTI encourage results reporting by its partners, they might serve as a framework for fostering better outcome-management in the public sector, no less than in the nonprofit one.

Four members of the SPEA Capstone Group were assigned to assess AIHTI's evaluation process and offer recommendations for improvement using the CN 3.0 framework. They used a modified version of the Charity Navigator Research Tool (the "Mock-Up") to examine how much results reporting the 17 host sites are currently doing and the extent to which they could be expected to do more.

This examination showed that the CN 3.0 framework can provide data necessary to evaluate charities with a wide spectrum of missions and goals. It also identifies recommendations for AIHTI on how to use the CN 3.0 framework to better evaluate the programs its members deliver at each of their sites.

Methodology

The Mock-Up was built on the CNRT. It includes questions on each of the four Elements, with some changes as recommended by the Capstone Group. In addition, the Mock-Up collected basic information about each program: how many services it offers, how many communities it serves, what partnerships it has with other organizations, and what its budget

is. (See Appendix A.) The Mock-Up also asks questions about the alignment between the mission of the organization (in AIHTI's case, improved physical activity level and increased health knowledge) and the programs being delivered at the 17 host sites.

The AIHTI case study focuses solely on the programs run by AmeriCorps members at each host site. Since no AmeriCorps resources are used for other programs (though the host site may conduct other activities), this had the effect of invalidating the CNRT's questions about the alignment of resources and mission. For assessing its results reporting practices, each of the 17 host-site programs was essentially treated as though it were a separate charity whose resources – from AIHTI or other sources – were completely dedicated to accomplishing its goals.

Since AIHTI requires that its members submit quarterly evaluations of their programs, the Capstone Group members examined them to determine whether they provided sufficient data to analyze each program according to the Mock-Up. They determined that the second and fourth quarter reports would be most useful, since they describe the program being evaluated, the process by which the evaluation is to take place, the indicators to be collected and the process for doing so, and how the results reported are to be used in the future.

These reports are completed in a format set by AmeriCorps, which assures that AIHTI members are providing quantitative data, along with program success stories and other anecdotal material. The first section of each report consists consists of demographic information about the program, such as the number of volunteers recruited, how many hours they served, and the number and type of populations served by the program. The second section is where the AmeriCorps members can provide an anecdote from their program and

attach documentation in the form of pictures, flyers, and handouts. The third section asks for information about the performance measures of the programs. Members must report outputs, anticipated results, indicators, information about their collection instruments, and so forth. The quarterly reports conclude with a section on the outcomes of the program.

By using these reports, the Capstone Group members felt they could best obtain the information needed to complete the Mock-Up (as well as see what AmeriCorps members were providing to AIHTI leadership and the Corporation for National and Community Service). Although CN 3.0 aims to evaluate information provided on charities' websites, looking at the websites of the 17 host sites would not produce sufficient information about the programs in which AmeriCorps members were serving (and which were to be evaluated).

Element One: Logic, Results, and Measures

The Logic, Results, and Measures element seeks to assess how well a charity has identified and reported the causal link between its activities and pre-defined outputs and outcomes.

The first test of this Element asks about the plausibility of the causal logic used. During their training sessions, AmeriCorps members are taught the theoretical concept of logic models and their utility in demonstrating the causal link between an organization's goals and programs. However, after examining evaluation documentation and discussing them with AIHTI leadership, it appeared that none of the AIHTI-supported programs were utilizing logic models as part of their evaluation efforts. Although the vast majority of the programs pass the initial causal logic question of having a mission statement with clear goals and objectives, they did not pass the subsequent question about aligning the mission statement with the services provided

by the program. The Capstone Group members did not look for evidence of strategic planning, since this process would be more relevant to the host-site, not an individual program within it.

The second test centers on the concept of dosage, or how much of a program is required to bring about the desired results. Although there is a clear description of the programs conducted by each AmeriCorps member, none of the 17 AIHTI programs report on the amount of time or effort required to impact participants.

Dosage is especially relevant to consider when conducting pre- and post-tests, or post-test only evaluations. An important strength of the pre- and post-test evaluation is it allows the evaluator to eliminate some outside influences by comparing one individual's scores at multiple points in time (Interview Dr. Hightower-King). However, AmeriCorps members sometimes make the pre-tests too easy leaving little room for improvement and many programs face serious issues with attrition between the pre- and post-tests. AmeriCorps members must know how many sessions a participant should attend in order to receive the full value of the program's education (King, 2012). This question of dosage is currently unaddressed in the AIHTI evaluation process, but should be a very important component to determining program success.

The third test focuses on whether reasonable evidence exists for the program's logic model. This test is seeking evidence of any academic research, case studies or other similar impact evaluations that justify the program's purpose and intended outcomes. None of the 17 AIHTI programs include this information in the quarterly evaluation reports.

This test also looks to see if an organization describes the need for its programs. While the 17 AIHTI programs at the host sites fail to describe the need for their specific programs,

AIHTI does cite the overall problem obesity in Indiana and offer statistical information regarding health problems in its grant applications and on its blog.

The fourth test examines the measures and methodology used to evaluate an organization's programs. It looks for whether an organization has identified short-term and long-term outcomes, indicators for those outcomes, and a collection strategy to gather data. In the quarterly evaluation reports, each AmeriCorps member is asked to provide intermediate and long-term outcomes, an indicator that will show progress in achieving the stated outcomes, and the instrument that will be used to gather the information on the indicator. Therefore, all 17 AIHTI programs at the host sites pass this test. The only information AmeriCorps members do not provide is a strategic plan regarding their respective AIHTI programs, but as noted previously, strategic planning is more a responsibility of AIHTI or the host sites on an organizational level than individual AmeriCorps members or programs.

Element Two: Independent Evaluations and Standards and Certification Mechanisms

Element Two focuses on whether an organization conducts independent evaluations of its programs. This Element also examines if an organization participates in any standards or certification mechanisms that require outcome measurement and reporting.

The first test of the Mock-Up is looks for a policy regarding evaluations. AIHTI requires each AmeriCorps member to submit quarterly reports with output and outcome data. Therefore, all 17 AHITI programs at the host sites pass this test. These quarterly progress reports on the AIHTI programs are a step in the right direction for evaluation, but unfortunately, none of the 17 AHITI programs have a policy that require an *independent* evaluation. Since one of the goals of AIHTI is to provide AmeriCorps members with the

experience of creating and implementing a program evaluation, each one is required to evaluate at least one of the programs he or she runs. Hence, none of the AIHTI programs can have a truly independent evaluation, unless done outside the AmeriCorps program.

The second test asks whether the results of the evaluations are published. Neither the host sites nor AIHTI publish the results of the evaluations conducted by AmeriCorps members on their websites. However, AmeriCorps members do have to report these results to AIHTI in their quarterly reports.

The third test asks if there are reports on corrective actions undertaken as a result of the evaluations. None of the host sites passes this test. This result should not be surprising, since outcome evaluations are conducted at the end of an AmeriCorps member's term and the transition process from one AmeriCorps member to the next at each host site does not necessarily include reviewing the results of previous evaluations. As such, any potential corrective actions that should be taken tend to be overlooked.

The final part of Element Two pertains to a charity's participation in outcome-oriented standards and certification mechanisms. Typically, these accrediting groups focus on the organizational level, not on specific programs. Because of this, the Capstone Group members feel this part of Element Two is not applicable for this case study, though to the extent standards and certification mechanisms encompassed particular programs, they deserve to be noted.

Element Three: Constituent Voice

"Constituent Voice" is "the practice of ensuring that the views of all relevant constituents are seriously taken into account in planning, monitoring, assessing, reporting and

learning from an organization's work" (Keystone Accountability, 2012). The third Element of Charity Navigator 3.0 attempts to assess how well a charity listens to its primary constituents—the beneficiaries of its services. The feedback it collects should go beyond narratives that could be used for marketing purposes, but instead involve systematic collection of data relevant to understanding the impact the charity is (or is not) having. This type of feedback may "provide organization leaders with information for strengthening performance, maximizing critical management values, such as efficiency, improving organizational learning and assessing the overall effectiveness and impact of a charity's work" (Bonbright, Campbell & Nguyen, 2009, p. 16). After reviewing each AIHTI host-site program, the members of the Capstone Group found no evidence that AmeriCorps members were collecting collect substantive feedback from program participants.

The initial Constituent Voice test concerns an organization's expression of the value of accountability and accessibility to primary constituents. While AIHTI and host sites display contact information primary constituents could potentially use to volunteer their feedback, the 17 programs do not explicitly solicit substantive feedback.

The second test considers the systems an organization has in place to rigorously collect primary constituent feedback. AmeriCorps members use a question bank, developed and provided by AIHTI leadership, from which to select evaluation questions. Though AIHTI's activity logs and pre- and post-tests attempt to measure changes in the activity and knowledge of program participants, they do not appear to ask specifically for constituent feedback. In their quarterly reports, AmeriCorps members are asked to record a "great story," which gives

them an opportunity to talk about program participants. However, these stories are used for marketing and are thus do not meet the standard for Constituent Voice.

Since AmeriCorps members do not currently collect constituent feedback, the third and fourth tests concerning analysis and utilization of feedback respectively are not applicable.

Recommendations

By using the CN 3.0 framework, AIHTI can improve its evaluation process substantially. To that end, the members of the Capstone Group who conducted this case study offer the following recommendations:

Element One

- Teach AmeriCorps members how to apply logic models as part of their training on program design, as program design is a critical component to any successful evaluation (King, 2012; McDavid & Hawthorn, 2006).
- Since the current AmeriCorps evaluation process does not include quantifying dosage, pair collected post-test results with participant attendance rates to more accurately estimate a threshold attendance level for each program to be evaluated (King, 2012).
- Move to a post-test only model, where AmeriCorps members focus on dosage required (King, 2012). A post-test only model makes no assumptions about participant knowledge of the subject that is being taught. Instead, it seeks to ascertain where participants wind up, depending on how much time or effort they put into the program. It would help provide a response to the question asked in the second test of Element One and which is currently not being

addressed in AIHTI's evaluation process: "How much action is required to produce the desired effects?" Once AmeriCorps members can determine the amount of participation needed to achieve a desired result, they can look at groups of participants who attain various outcomes and adjust programs accordingly to produce greater impact.

Element Two

- Engage evaluation experts from Indiana University's School of Public and Environmental Affairs and School of Health, Physical Education, and Recreation to aid AmeriCorps members in creating evaluation tools. This outside expertise will offer additional support to improve the quality of the evaluation tools used to measure intended outcomes and give greater credibility to the results.
- Have someone in addition to the AmeriCorps member assess the data collected via the evaluation. Under current procedures, each AmeriCorps member is solely responsible for creating and reviewing evaluation instruments. While this provides valuable experience for the member, consistent with AIHTI objectives, it raises questions about how independent the evaluation is, since the member is also responsible for the program being assessed. Asking another person – including, perhaps, another AmeriCorps member – who has not been involved with the program to review evaluation methods and results would make the process more consistent with Element Two.
- Require AmeriCorps members to identify corrective actions to improve programs based on evaluations. After analyzing the results, AmeriCorps members should

include in their reports lists of the improvements that should be made. That would enable the AmeriCorps members that follow them in the same programs to build on previous years' evaluations by implementing the recommended steps and determining how they affect outcomes.

Element Three

- Show commitment to the opinion and voice of program participants by adding age-appropriate, qualitative questions concerning satisfaction, barriers to attendance, utility, how to improve the service, etc., to the evaluation instruments. For comparability, the same questions could be asked across all programs. AIHTI can add these questions to the question banks AmeriCorps members use to develop evaluation instruments.
- Once collected, AIHTI should publish the aggregate results of feedback from host sites on its website (Charity Navigator 2012). Providing the results of feedback will allow future AmeriCorps applicants to have a better understanding of the effectiveness of AIHTI programs.
- Report results of feedback to primary constituents (Charity Navigator, 2012). Reporting back has the potential to create “learning relationships with [an organization’s] primary constituents” (Bonbright et al., 2009, p. 30). Even a young child may feel a sense of satisfaction and efficacy when a program is changed because of something he or she said.

Conclusion

Based on this case study, AIHTI can alter and improve its evaluation process by using the Charity Navigator 3.0 concepts. Despite the constraints of AmeriCorps programs, including annual turnover of members and participant attrition, and the wide range of host sites, employing CN 3.0 Elements not only seems feasible (although pilot-testing would be useful), but also should not create a financial burden for AIHTI, which does not have additional resources to spend on improving its evaluation process. However, AIHTI will need to provide more guidance to AmeriCorps members as they develop programs in order for them to be aligned with the evaluation criteria. Providing guidance on program design will allow AIHTI to continue giving members the experience of creating and evaluating a program, but will also help foster more effective programs and improve evaluations. Indiana University's School of Public and Environmental Affairs and School of Health, Physical Education, and Recreation also have faculty and graduate students to assist AIHTI with this.

The case study also showed that the host sites currently employ few of the concepts incorporated in CN 3.0 in their AIHTI-supported programs. Examining quarterly reports covering a longer period might have affected this conclusion, but the major challenge is the long-existing AmeriCorps goal of giving its members experience in program development and evaluation design. But the Capstone Group members who did this case study do not regard that as an insuperable obstacle. For example, while it may never be possible for AIHTI programs to have completely independent evaluations, the recommendations above suggest how they can be made more objective and useful than they current are and as CN 3.0 aims to do.

Finally, it should be noted that AIHTI host sites are not part of the same cause areas as the charities that were researched by the rest of the SPEA Capstone Group. The Charity Navigator research was conducted on a sample of charities providing Homeless Services, Social Services, and Children's and Family Services. Additionally, the case study treated each host site's AIHTI program as if it were a charity by itself.

Nonetheless, by mapping the Mock-Up version of the CNRT to a set of programs outside the realm of those in Charity Navigator's data base, this case study has demonstrated that the CN 3.0 framework is sufficiently robust to be applicable in the context of a public-private partnership, such as the one established between AIHTI and its 17 host sites and help it by applying a rigorous rating approach to assessing its outcomes.

Chapter Eleven

Recommendations and Conclusions

CN 3.0 is a promising tool for enabling Charity Navigator to continue its efforts to increase donor effectiveness by driving donations to charities with the strongest results reporting. For the Tool to be worthwhile and to reduce variation among raters, the Capstone Group recommends that Charity Navigator take steps to reduce subjectivity in the tests for each of the Elements, provide explicit instructions and frameworks to ensure common understanding among raters, clarify the language and terminology for each of the Elements, and utilize different structures of questions to provide more precise indications of charities' results reporting efforts.

To illustrate how these changes could be made, the SPEA Capstone Group revised the Charity Navigator Research Tool, changing question language, structure, and format to provide clarity and understanding to both raters and charities. To show what future raters would see going through the rating process, the resultant "Mock-Up" for a new tool is presented in a web interface format similar to the current CNRT. (See Appendix A, Mock-Ups of Revised and Original Tools.). To enable Charity Navigator or future researchers to determine patterns of concern or opportunities for advancement in the non-profit sector, the SPEA Capstone Group also added another section to record a charity's demographic or geographic characteristics, such as size, location, and years of existence.

For Element One (Logic, Results, and Measures), each test was disaggregated into multiple components. For instance, the concept of causal logic was divided into mission,

objectives, and goals. Also, the Capstone Group proposed an additional question, asking raters to evaluate the strengths and weaknesses of each test, so Charity Navigator can continue to refine them based upon rater feedback. It also changed the structure of some tests, making some questions open-ended and providing more answer choices than the original binary format.

For Element Two (Independent Evaluations and Standards and Certification Mechanisms), the SPEA Capstone Group disaggregated the question about independent evaluations into multiple questions, first asking if the charity published any evaluation results and subsequently asking about the type of evaluation published. This test also gives more answer choices; the Capstone Group recommends creating a checklist of types of evaluations, including independent, collaborative, and internal. Another recommendation for Element Two is to clarify the Research Guidelines to define and provide examples of outcome-oriented standards and certification mechanisms to guide the raters. Finally, an additional question to gauge the transparency of a charity's performance or awards from standards and certification mechanisms is recommended.

For Element Three (Constituent Voice), the SPEA Capstone Group re-ordered the tests by applying the four-component framework of values, systems, analysis, and utilization. It also recommended disaggregating tests, for example, by dividing the question asking if a charity rigorously collects feedback into two parts: first, if the charity collects feedback and second, how rigorous that system is. Also, the SPEA Capstone Group advises changing the threshold or screening role of the first test by using a rating scale, ranging from "Not at all Rigorous" to "Very Rigorous" with "Not Applicable" as an answer choice, rather than the current binary

scale. Another recommendation is to provide definitions for technical terms, like “primary constituents,” that link to a glossary website. Finally, three additional tests were created to incorporate aspects of Constituent Voice currently not included in the CNRT.

For Element Four (Alignment of Mission and Resources), the SPEA Capstone Group edited the CNRT’s Compliance Footprint questions in the Mock-Up, since the specific questions for Alignment of Mission and Resources were not available when the initial survey of 90 charities was begun. For example, the Compliance Footprint tests were revised from a 25 percentage-point scale to a 10 percentage-point scale.

Due to some uncertainty and concerns about the new framework among charities, the Capstone Group recommends an active effort to communicate with them about the importance and relevance of the CN 3.0 framework. Charity Navigator can also work to assist and educate charities through the CN forum during the transition to the roll-out of CN 3.0. Improving charity acceptance of this framework is essential if it is to accomplish its goal of advancing the nonprofit sector, since only through positive relationships and collaboration will Charity Navigator successfully affect the mindset of charities with regard to the value of better results reporting.

The Capstone Group’s research also shows that CN members are interested in outcome-oriented information when making decisions to donate and look to Charity Navigator to collect and judge that information. Since members do have concerns about the capacity of charities to provide the necessary data accurately (and at all), as well as a limited understanding of the constituent voice concept, Charity Navigator will need to try to educate them on the Element particularly as it begins utilizing CN 3.0.

Finally, the Capstone Group recommends that Charity Navigator seek out partnerships through AmeriCorps and other organizations nationwide to broaden the use of its 3.0 framework to other contexts besides its own ratings system, including public-private partnerships, to increase the likelihood that results reporting will become the norm for efficient charities generally in the future.

Works Cited

- AHITI Members. (2012). Personal Communication: Interview. Friday, April 6 2012.
- Allison, M & Kaye, J. (1997). "Chapter One: Introduction to Strategic Planning," in Allison, M. and Kaye, J. (Eds.), *Strategic Planning for Nonprofit Organizations*, 2nd ed. (2005). Hoboken, NJ: Wiley.
- "AmeriCorps Improving Health Throughout Indiana." (2012). Retrieved from <http://www.hper.indiana.edu/health/community/index.shtml>.
- Better Business Bureau's Wise Giving Alliance. (Accessed 4 February, 2012). Retrieved from <http://www.bbb.org/us/Charity-Standards/>.
- Bonbright, D., Campbell, D., & Nguyen, L. (2009). *The 21st Century Potential of Constituency Voice*. Retrieved from http://www.keystoneaccountability.org/sites/default/files/Constituency-Report-final-27APRIL2009_0.pdf.
- Bradburn, Norman, Sudman, Seymour, and Wansink, Brian. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design- for Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco: Jossey-Bass.
- Charities Review Council. (2010). *Governance standards*. (Accessed 4 February, 2012). Retrieved from http://www.smartgivers.org/governance_standards.html.
- Charity Navigator. (2012). *Charity Navigator 3.0 2012 Research Guidelines*. Communication to SPEA Capstone Group. February 15, 2012.
- Charity Navigator. (n.d.). *Home*. (Accessed March 2012). Retrieved from <http://www.charitynavigator.org>.
- Charity Navigator. (n.d.). *Mission*. (Accessed March 2012). Retrieved from <http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=17>.
- Charity Navigator. (n.d.). *About Us*. (Accessed March 2012). Retrieved from <http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=628>.
- Charity Navigator. (n.d.). *Methodology*. (Accessed March 26, 2012). Retrieved from <http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=33>.
- Charity Watch. (n.d.) *Criteria*. (Accessed March 2012). Retrieved from <http://charitywatch.org/criteria.html>.

- Charting Impact. (Accessed 20 February, 2012). Retrieved from <http://www.chartingimpact.org/about/five-questions/>.
- Cooperman, Erin. (2012). Personal Communication: Interview. Thursdays, January - April 2012.
- Ebrahim, A. (2010). "Chapter 4: The Many Faces of Nonprofit Accountability," in Renz, D. O. and Associates, Eds., *The Jossey-Bass Handbook of Nonprofit Leadership and Management, 3rd Edition*. San Francisco: Jossey-Bass.
- Farris, R., Finklestein, E., Khavjou, Olga., & Will, J. (2007). "Beyond Effectiveness: Evaluating Public Health Impact," *American Journal of Public Health*, 97 (4) 641-647.
- The Fundraising School (TFRS). (2011). *Principles and Techniques of Fund Raising* Indianapolis: The Center on Philanthropy at Indiana University.
- GiveWell. (n.d.) *Criteria*. (Accessed 5 February, 2012). Retrieved from <http://givewell.org/international/technical/criteria/impact>.
- GiveWell. (n.d.) *Process Overview*. (Accessed 5 February, 2012). Retrieved from <http://givewell.org/united-states/process>.
- Gronbjerg, Kirsten. (2012). Personal Communication: Interview. February 3, 2012.
- Gronbjerg, K. A. (2001). "The U.S. Nonprofit Human Service Sector: A Creeping Revolution," *Nonprofit and Voluntary Sector Quarterly*, 30 (2), 279-97.
- Gugerty, M. K. & Prakash, A., eds. (2010). *Voluntary Regulation of NGOs and Nonprofits: An Accountability Club Framework*. Cambridge, UK: Cambridge University Press.
- Guidestar. (n.d.) *Take Action*. (Accessed 20 February, 2012). Retrieved from <http://www.guidestar.org/rxg/give-to-charity/questions-to-ask-a-nonprofit-before-investing-in-it.aspx>.
- Hightower-King, Mindy. (2012). Personal Communication: Interview. March 7, 2012.
- Hopkins, B. (2009). "Chapter 21: Watchdogs on the Prowl," in Hopkins, B., *Starting and Managing a Nonprofit Organization, 5th Edition*. Hoboken, NJ: John Wiley & Sons.
- Huitt, B., Hummel, J. & Kaeck, D. (1999). "Internal and External Validity: General Issues," Retrieved from <http://www.edpsycinteractive.org/topics/intro/valdgn.html>.

- Independent Sector. (n.d.) *Compendium of Standards, Codes, and Principles of Nonprofit and Philanthropic Organizations*. (Accessed March 2012). Retrieved from http://www.independentsector.org/compendium_of_standards.
- Kearns, K. P. (2001). "Chapter 31: Accountability Concepts and Controversies," in Ott, J.S., Ed., *Understanding Nonprofit Organizations: Governance, Leadership, and Management*. Boulder, CO: Westview Press.
- Keystone Accountability. (2012) Glossary. Retrieved from <http://www.keystoneaccountability.org/glossary>.
- Light, P. C. (2002). *Pathways to Excellence*. Washington, D.C: Brookings Institution Press.
- Lucas, J. (2001). "Vision: Having a Mission Statement Means We Know Who We Are," in *Fatal Illusions: Shredding a Dozen Unrealities that can Keep Your Organization from Success*. Boston: Quintessential Books.
- McDavid, J. C., & Hawthorn, L. R.L. (2006). *Program Evaluation & Performance Measurement: An Introduction to Practice*. Thousand Oaks, CA: Sage Publications, Inc.
- Migliore, R. et al. (1995). "Defining Organizational Purpose," in Allison, M. and Kaye, J. (Eds.), *Strategic Planning for Nonprofit Organizations*, 2nd ed. (2005). Hoboken, NJ: Wiley.
- Miller, Will. (2012). Personal Communication: Interview. February 24, 2012.
- Mitchell, R. (2010) "It's time to measure nonprofit effectiveness," *Philanthromax*. Retrieved from <http://philanthromax.com/blogs/rob-mitchell/it%E2%80%99s-time-measure-nonprofit-effectiveness>.
- National Council of Nonprofit Associations and National Human Services Assembly. (2005). *Rating the Raters: An Assessment of Organizations and Publications that Rank/Rate Charitable Nonprofit Organizations*. Washington, D.C. Retrieved from <http://www.nationalassembly.org/Publications/documents/ratingtheraters.pdf>.
- Ott, J.S. (2001). "Part 10: Accountability and Evaluation," in Ott, J.S., Ed., *Understanding Nonprofit Organizations: Governance, Leadership, and Management*. Boulder, CO: Westview Press.
- Reed, E., & Morariu, J. (2010). "State of Evaluation: Evaluation Practice and Capacity in the Nonprofit Sector," *Innovation Network*. Retrieved from http://www.innonet.org/client_docs/innonet-state-of-evaluation-2010.pdf.
- Rodriguez-Campos, L. (2005). *Collaborative Evaluations: A step-by-step Model for the Evaluator*. Charleston, SC: BookSurge Publishing.

SAGE Eldercare. (n.d.). *Summary of Case for Support*. Retrieved from:
<http://www.sageeldercare.org/about-sage/documents/CaseStatement09.pdf>.

Sargeant, Adrian and Shang, J. (2010). *Fundraising Principles and Practice*. San Francisco, California: Jossey-Bass.

Standards for Excellence Institute. (n.d.) *The Standards for Excellence Code*. (Accessed 20 February, 2012). Retrieved from
<http://www.standardsforexcellenceinstitute.org/dnn/TheCode.aspx>.