



CN 3.0 Concept Note

The Evaluation of Reporting Results

December 2012

Table of Contents

Step One: Report Results	1
What do we mean when we refer to a charity's results?	2
A Developmental Approach	3
The Five CN 3.0 Rating Elements	5
Rating Element One: Alignment of Mission, Solicitations and Resources	6
Rating Element Two: Results Logic and Measures	7
Rating Element Three: Validators	10
Rating Element Four: Constituent Voice	11
Rating Element Five: Published Evaluation Reports	13
The Problem of Cheaters	14
Conclusion: The Way Forward	15
Glossary	17

Step One: Report Results

In order to generate financial support for the most effective organizations it is necessary, among other things, for organizations to publicly report their mission-related results. This statement is an uncontested prerequisite for public companies, at least when it comes to their mission-related “bottom line” results. For most public charities, with respect to their mission-related results, it is at best an aspiration.

The state of results reporting by America’s charities makes it very difficult for conscientious donors to decide where best to invest their charitable dollars in order to have the most beneficial impact on those who are most in need. The current norm of charity reporting, as represented by websites and annual reports, is little more than promotional marketing. It is not possible today to make even preliminary assessments of the effectiveness of the overwhelming majority of America’s charities on the basis of what those charities publish about their results. Charity Navigator’s newest rating dimension – CN 3.0 – seeks to change this by recognizing charities that do a better job of reporting their results. This provides effective charities with a particular opportunity to demonstrate their effectiveness by reporting their results. As more charities report their results, givers and social investors will have more and better information to inform their decisions.

In the future, when more results information is available, Charity Navigator will assess those reported results. But before it is possible to evaluate results, it is necessary to shift the paradigm of charity reporting from the current norm of selectively reporting, such as storytelling or case studies that may not be representative of overall performance, to reporting on demonstrably important measures, and showing how the organization learns and improves based on those measures. CN 3.0 seeks to contribute to such a paradigm shift.

What do we mean when we refer to a charity's results?

When we refer to a charity's results, we can think of this as shorthand for what we expect them to achieve with our donations. But many charitable results that we most care about are difficult to measure. The results that are easier to measure are often not the most important results. To take an example from the current debates about poverty alleviation, measuring repayment rates for microloans is important, but they will not tell you what is most important – if those microloans are reducing poverty.

Time is also a significant factor in understanding charity results. **When** you measure can shape your evidence as much as **what** you measure. Some of the most important results do not occur until much later than the actions that contribute to them. The influence that a wise relative, a great teacher, or a skilled counselor may have on a young person may not show up for years.

Then there is the conundrum of causality. When can a charity legitimately claim that it caused a particular result? No single organization is ever the sole cause of a social outcome. By their natures, social outcomes are the products of multiple factors. The microloans to a business entrepreneur who succeeded to build up a growing business surely contributed to her success. But can we say they caused her success? What is reasonable for a charity to claim about any given result? What measurement systems should a charity have in order to make what claims about results?

These are just some of the complexities in measuring and comparing charitable results that we bear in mind as we set CN 3.0 rating criteria. At the same time, we believe that effective charities manage their performance and thereby know and act on their results. These high performing organizations use early results as milestones to the important longer-term results. Accordingly, the CN 3.0 approach to rating charity results is focused on the ways charities come to know and use their results.

A Developmental Approach

CN 3.0 is part of a trend towards better performance and results measurement among America's charities. The bar for charity results measurement and reporting is rising. CN 3.0 rating criteria are in line with the leading practices of those charities that are doing the most to raise this bar. At the same time, Charity Navigator recognizes that the majority of charities are in the early stages of measuring and reporting results and do not yet have the necessary systems in place to meet the CN 3.0 rating criteria. For this reason, we are taking a developmental approach to the introduction of CN 3.0, by which we mean that we will engage with charities with the intention of encouraging and incentivizing progressive improvement in results measurement and reporting practices. Charity Navigator has worked with a panel of expert advisors – including leaders from several charities – in designing this process and has incorporated much of the feedback. As with other elements of the Charity Navigator rating system, Charity Navigator provides a way for charities to comment on any element of the Charity Navigator rating of their organization.

As charities demonstrate improved results reporting practices, the CN 3.0 rating criteria will evolve to keep raising the bar of reporting requirements.

The developmental approach is realized through a number of interrelated parts:

- 1. Five simple transparency-based rating criteria:** Charity Navigator needs clear rating criteria that are applied consistently across charities and that will enable meaningful comparisons across the rated charities. The five rating elements discussed below focus squarely on the way charities report results and revolve around three foundational aspects to valid measurement of charity results – standard definitions of measures, rigor in measurement practices, and external validation.

They also recognize the degree of difficulty associated with different kinds of results evidence from easier-to-measure indicators of execution (outputs) to harder-to-measure indicators of change (outcomes and impacts).¹ Charity Navigator 3.0 rating criteria will give some credit to charities with performance management and measurement systems that generate less important evidence of social value creation (outputs) but are easier to measure. However, our rating criteria are designed to create incentives for using measures of outcomes since these have been shown to be necessary to being an effective, high-performing organization.

¹ For one of the more thoughtful recent efforts to classify charity results see the Bill & Melinda Gates Foundation's *Actionable Measurement Guide*. For our definitions of results evidence, see the Glossary at the end of this concept note.

2. **Charity preview and input process:** Charity Navigator is developing a process wherein the charity has an opportunity at the very beginning of our review of the organization to direct us to relevant web pages and other supportive materials. In this way charities can assist us in completing the review as well as increasing the likelihood that we will consider all relevant information.
3. **One-year grace period:** For the first round of CN 3.0 ratings, charities will be able to provide the evidence required to meet CN 3.0 rating criteria directly to Charity Navigator. For a charity to continue to get credit towards CN 3.0 ratings after its first year, the evidence supporting that rating must be available to the general public through the charity's website.
4. **Charity feedback:** We will continue in diverse ways to solicit charities' suggestions and concerns regarding CN 3.0 as well as what may or may not be appropriate metrics that are unique to different cause areas. To the maximum extent possible we will cultivate public discussion of CN 3.0. Our methodology will be available in full for charities to apply themselves to their measurement and reporting practices. With this information they will be able to identify and assess the relevance of the existing results reporting practices that are recognized by CN 3.0 ratings and/or that can be used to meet CN 3.0 rating criteria.
5. **Reporting results resources online:** If additional funding is available, Charity Navigator's consulting partner, Keystone Accountability, will create an affiliated but separate website with resources and tools for reporting results. The website will enable charity staff to:
 - a. Access measurement resources and tools directly as well as from the best existing knowledge repositories such as the Foundation Center's TRASI website. Illustrations of practices associated with top CN 3.0 ratings will be highlighted.
 - b. Share questions and experiences, engage with each other and outcome measurement experts.
 - c. Identify measurement and evaluation service providers. Charities will be able to identify consultants that are viewed to be effective by at least some external validator, and that will allow comparisons across providers by like-to-like characteristics. Charity clients of these providers will rate providers performance and these ratings will be published on the website as part of provider profiles, similar to eBay's buyer ratings of sellers.

The discussion of the CN 3.0 rating elements that follows provides examples of the kinds of tools and resources that could become available through a Keystone Accountability online portal cataloguing results reporting resources.

The Five CN 3.0 Rating Elements

This first version of CN 3.0 is designed to improve the practices charities use for results reporting by enabling Charity Navigator analysts and others to perform independent reviews of published evidence of results. The five rating elements in CN 3.0 have been selected specifically to establish three essential preconditions for effective arms-length review and analysis: standard definitions of measures (enabling data comparability), rigor in measurement practices, and external validation of results evidence.

Charity Navigator understands that it will take several years or more for most charities to publish evidence of their results in ways that are comparable across organizations and can be relied on as accurate because the data are independently validated. These five rating elements will encourage and incentivize charities to report results in ways that move us purposefully toward a day in the not-too-distant future when we can meaningfully compare charities on the basis of their results.

We do not anticipate resolving the weighting of the rating score system we will utilize, however, until we have gathered and posted data on our web site for all the charities in our database. Hence, it will require several years of research, going cause area by cause area, before we settle on the eventual scoring system. Meanwhile, as they are finalized, we will post the rating elements we establish for each cause area and then provide this information on the profile page of each charity within that cause area.

Rating Element One: Alignment of Mission, Solicitations and Resources

This element follows a “truth in advertising” principle, and operates as a “knock out test” in CN 3.0. It recognizes that a charity cannot satisfactorily meet its results reporting requirements if it creates an impression of its mission, work and results that are not in line with how it spends its finance and staff time. In addition, this element identifies the charity’s relevant program areas. This information is critically necessary as a first step to determine how many separate reviews of the other four elements need to be completed for each charity. Ultimately, when we begin to rate charities, this information will also impact the proportionate scoring as we rate results reporting data. For example, if the charity only has results logic and measures for one of three program areas, the rating will reflect that limitation. In addition, it is likely that an organization will not specify each of its program areas covered when it proceeds to answer the elements that follow. In that instance, our analyst staff will fill in this information for the largest program area and leave the other program areas blank. In this way we would like to provide the organization with some “credit” for reporting results but also encourage them to be more precise in identifying the relevant program areas in the future.

Charity Navigator applies a simple pass/fail test in assessing this element. If a charity meets the test, Charity Navigator goes on to assess the other four rating elements. If the charity fails Rating Element One, it cannot receive credit under CN 3.0. It will be reported that the charity failed this first fundamental element and the reviewer will go no further. The review is now complete.

The standard used for this element is a variation of the common law reasonable person standard. Applied here it reads, “After a review of the charity’s funding solicitation materials, (which must be on the donate page of its website, its public financial reports, or other materials sent to us) a reasonable person would conclude that the way the charity allocates its resources would ordinarily be appropriate to meet its solicitation claims.” Public financial reports may only be available for the previous financial year, but if even they are not available we will give charities the opportunity to submit recent reports directly to Charity Navigator. The specific data points to make this assessment will include the breakdown of program expenses as reflected in the charity’s annual 990 submission to the IRS as contrasted with its public pronouncements, solicitation requests, information provided on its web site as to what programs it provides.

Rating Element Two: Results Logic and Measures

This element looks for a set of clear statements indicating how and why the charities activities are intended to lead to pre-defined outputs and outcomes. These statements must also cover how results-related data actually collected by the charity are meant to mark progress against those results.

The rubric for this will be four yes/no questions. Illustrations of standards and applications of these tests will be published on the results reporting resources online portal.

1. Is the organization's causal logic plausible?

This simple test will ask if a reasonable person would find the charity's causal logic, as stated, to be plausible. By plausible we mean possible or likely under normal circumstances. If a reasonable person would conclude the charity's change logic is not plausible, then that causal logic fails the test.

Let's take the example of a job opportunities program. The logic statement for this hypothetical program is, "Our job classes provide our participants with the skills that they need to secure and hold a job." On the face of it, this is clearly plausible. On the other hand, if the logic statement were, "Our lucky charm bracelets provide our participants with all that they need to secure and hold a job," then the test would not be met.

2. Is there some indication of how much of the action is required to produce what effects?

If we were talking about medicine we would call this the dosage. This idea can be applied to all social interventions. No charity argues that any measure of exposure to its activities is going to produce the intended result. This test looks to see if the charity has indicated what level of activity is required. Charity Navigator recognizes that the value of knowing and measuring "how much" varies considerably by aims and activities. For example, spending time assessing measures of "how much" in awareness raising or advocacy or research may be more effort than it is worth. In these cases, this factor may not be weighted in the analysis or weighted to a lesser degree.²

To return to the job employment opportunities example, the statement ("Our job classes provide our participants with the skills that they need to secure and hold a job.") does tell us the result (secure and hold a job), but does not tell us how many or how long classes are required. As stated, this statement is insufficient to pass this test. In this case we look further to see if indications of quantity of classes are

² It is also reasonable to expect that charities may determine that the appropriate "dosage" may change; this test will not penalize charities that make such changes.

introduced elsewhere. For example, the organization might note that, “Completing four course modules over six months, and successfully demonstrating mastery of the material, is the required minimum participation to complete the basic program.” This statement would pass this test.

3. Is there some indication that the stated causal logic is based on reasonable evidence (i.e., references to other programs that have demonstrated results following the same logic)?

Here we look beyond the logic and the level of activity to see if the report includes reference to evidence that these actions have in fact been demonstrated to produce the intended results. References could be evidence from other organizations or from past practice by the charity itself.

In the hypothetical job opportunities program, the following statement would meet this test: “Over the past 5 years, 80 percent of our participants who pass four job courses within six months obtain employment within 6 weeks of graduating. For graduates of more than one year ago, 70 percent of these are still employed one year later.” Note that the evidence needs to speak to both aspects of the intended result – obtaining and retaining a job.

4. Are there specified measures (indicators) for outputs and outcomes to be collected and a plan to do so?

What we are looking for here is a simple plan for data collection of specified indicators that it is reasonable to assume will capture the intended result. In other words, does the charity state what success looks like and how it will collect the data to know if it has been achieved?

For our hypothetical employment opportunities program we would expect to see something like: “Our graduates inform us when they obtain a job. We survey both our graduates and the employers twice a year to monitor their status. We maintain client employment histories for seven years.” This represents a data collection plan and a clear specified indicator (employment status).

We are not assessing the quality of the data collection plan. In this example, having two sources of information for the same outcome indicates a strong plan. But if it were less compelling – for example, based on graduate surveys only – then it would still pass the test.

To pass the test both specified indicators and a plan for collection must be present.

After applying the four tests, we determine how much of the overall activities of the charity are covered by the rating element. We call this coverage the rating element’s footprint. For example, our research may lead us to conclude that a charity’s published logic model and measurement plan should get a 100 percent score on the

four rating tests, but since it only covered twenty percent of organization's mission it would only gain twenty percent of the total points available.

It bears noting that these four tests have much in common with a number of approaches and tools for planning and measurement that are increasingly favored among effective charities, such as theory of change, logical framework, results framework, logic model, constituent voice, most significant change, outcome mapping and others. This rating element distills from this family of planning and measurement tools what it most important for results reporting without taking a position on any of them.

Rating Element Three: Validators

This third element recognizes the important role that independent third parties can play in establishing the measurement credentials of a charity. They may be thought of as an analogy to independent or statutory financial auditors and are the closest thing we have to independent auditors to charity results. There are several different types and forms of these validators: standards bodies, codes of conduct, certification or accreditation regimes, shared measurement regimes, certified marketplaces, competitive open public award competitions, and charity analysts. Validators target many different dimensions of work – for example, health & safety, environmental impact, management operations, and more. CN 3.0 is only interested in those that have a focus on results measurement and reporting.

Organizations may receive credit for membership in good standing/participation in one or more validators now available to America’s charities. To receive CN 3.0 credit a charity must publicize its membership in and be a member in good standing (or equivalent) of the validator.

Independently, Charity Navigator is assessing validators for the likelihood that the validator provides assurance that the charity has adequate measurement and reporting practices. This assessment is based on two questions:

- 1. Does that validator explicitly and meaningfully cover results measurement and reporting?**
- 2. Is there a credible threat for non-compliance or a credible incentive for compliance?**

Charity Navigator will publish and continually update a registry of validators identified by Charity Navigator-rated charities. Ultimately (once we begin to rate charities on this element in a few years), the registry will explain how Charity Navigator scores each validator, and indicate how many CN 3.0 points are allocated for membership in a recognized validator. Validators will be encouraged to review their scores to make sure the Charity Navigator team has not missed anything.

Categories (and examples) of validators: accreditation agency (e.g., Council on Accreditation), standards body (e.g., InterAction Private Voluntary Organisation Standards), shared measurement system (e.g., Success Measures), charity analyst (e.g., Root Cause), and public competitive award (e.g., Social Innovation Fund).

Rating Element Four: Constituent Voice

This rating element assesses whether and how well a charity collects and publishes feedback from its primary constituents, variously referred to as beneficiaries, clients, end users or consumers. Many charities are careful to solicit and be responsive to feedback from their donors. When it comes to mission, however, the critically important constituents are the individuals who are meant to receive the ultimate benefits of the organization's work, or those (such as the parents of young children) who may act as proxies for that constituent.

Some have asked, "What about charities whose aims are to protect the environment or other species? What about charities dedicated to human beings who do not have a voice, such as future generations, prisoners, very young children, or those with diminished capacity?" The answer to this question begins with rating element two, in the charities' results logic. Whatever the nature of a charity's objectives, its activities will be seeking to change something about which some people are well positioned and able to provide important feedback. These are its primary constituents for the purpose of this rating element. There may be very rare exceptions to this, but Charity Navigator has not come across any yet. For the environmental charity, it might be biologists or policy makers. For a human rights charity advocating for political prisoners it might be a combination of researchers, journalists, policy makers and peers. For a charity working for very young children it might be parents, or local officials or policy makers. Or for each of these hypotheticals it could be some other group. To find the primary constituent, one must trace the organization's results logic.

Charity Navigator is researching the application of six yes/no tests to determine whether the organization is accountable, responsive and effective from the point of view of its primary constituents.

- 1. Does the charity publish feedback data from its primary constituents?**
- 2. Does the published feedback data include an explanation of how likely it is to be representative of all primary constituents?**
- 3. Does the data include an explanation of why the organization believes the feedback is frank and honest?**
- 4. Is that data presented in a way that shows changes over time going back at least one year?³**
- 5. Does the data include questions that speak to the organization's effectiveness?**
- 6. Does the organization report back to its primary constituents what it heard from them?**

³ Charity Navigator will be careful not to penalize organizations when they first report for not having prior year data.

Charity Navigator recognizes that organizations that publish feedback may be asking questions that do not generate much value. This is analogous to the general point that at this time CN 3.0 is assessing results reporting, not the results themselves. Once more charities are publishing feedback data, we will have a new informational basis from which to debate and raise the quality of that data. In the future, later versions of CN 3.0 will gradually raise the bar by applying stricter tests for Constituent Voice. In the meantime, organizations that are interested in good Constituent Voice practices may refer to the guidelines, resources and free tools available on the Keystone Accountability website.⁴

⁴ www.keystoneaccountability.org

Rating Element Five: Published Evaluation Reports

This fifth rating element recognizes the value of evaluation in understanding and enhancing an organization's ability to achieve its intended results.

In practice, evaluation methods applied by organizations vary widely in type, purpose, scope, design and implementation quality. Moreover, the value that organizations derive from "doing evaluation" is oftentimes determined by factors beyond the evaluation itself. This rating element does not try to directly address or resolve this complexity in ascertaining the differential values of particular evaluation practices or instances. It operates from the hypothesis that organizations that undertake and publish regular formal evaluation reports that consider the results of the majority of their program activities on a regular basis are more likely to understand and achieve their intended results than those organizations that do not. The next few years of applying the three yes/no tests to be applied in this element will provide a test of this hypothesis.

In this rating element, Charity Navigator is researching the application of the following three yes/no tests.

- 1. Does the charity publish evaluation reports that cover the results for its programs at least every 5 years?**
- 2. Are those evaluation reports based on recognized techniques to understand their results? Charity Navigator will publish a list of such techniques and keep it updated. Charities can make the case for specific techniques that they have found effective. The initial list of recognized techniques is: before and after measures; cost-to-benefit analysis, long-term follow up, studies with control/comparison groups, and randomized control trials.⁵**
- 3. Either in the evaluation report or subsequently, does the charity explain what, if anything, it is changing as a result of the findings in the evaluation report?**

⁵ For an up-to-date review of the state of the art in impact evaluation, see the InterAction Impact Evaluation Guidance Note series at <http://www.interaction.org/impact-evaluation-notes>.

The Problem of Cheaters

CN 3.0 has been designed to minimize the room for cheating (i.e., reporting of false results) in several ways. At least for the first few years, charities will not be rated on their results, but on the how they report those results. This puts the focus on the integrity of their measurement practices while avoiding penalizing organizations for what might be interpreted as poor results.

Reported results can be studied by a variety of third parties. We have incorporated formal third-party reviews through rating element three. In developing CN 3.0, we have worked with ten graduate schools interested in nonprofit performance and management. This productive set of interactions suggests that there is a large and active community of professors and students that will test the assumptions in CN 3.0 through rigorous research. In particular, we invite researchers to test our hypothesis that charities that measure and know their results are more likely to achieve results than charities that do not report results. Conversely, we anticipate that charities with poor results will not be interested in sharing this information publically, particularly when these results will be compared against charities that take results reporting seriously.

Finally, we will set severe penalties for charities that are caught cheating. Penalties are likely to include inclusion in our Donor Advisory system, which has a documented negative impact on a charity's fundraising.

Conclusion: The Way Forward

Why are we engaging in what is rightly characterized as a herculean effort?

Our answer has two parts. First, we think we have designed a process that will yield a meaningful and balanced overall rating model for charities that includes results reporting. Second, we believe that once in place, these ratings will channel funding to many higher performing charities that proudly publicly report on their results. This would motivate many more charities to become higher performing so they can produce better outcomes and also publicly report these achievements to gain higher ratings from Charity Navigator and secure more donor support.

With respect to the first part of the answer, we have designed an iterative, client- and constituent-centered approach to CN 3.0 development. This concept note indicates how we will apply an initial set of results reporting elements to one cause area among the 34 we consider in our database of charities. As should be evident from the cautious terminology we have attempted to utilize here, we anticipate that this effort will require a number of years of research and analysis before we can cover and customize our rating elements to fit the circumstances and evolutionary stage of the various types of charities we evaluate.

In summary, our approach going forward will:

- Select one cause area at a time, building a cumulative understanding of the specificities and appropriate results reporting elements of different cause areas.
- Defer the design of a scoring system until we have data on all the charities we evaluate (as we did with the CN 2.0 roll out).
- Ensure that, in the transition period to CN 3.0, we do not have an “apples and oranges” approach to rating some charities in one way (CN 2.0) and others in another way (CN 3.0).
- Dialogue continuously with experts and the charity community to assure we maximize their feedback and expertise.
- Check in with our primary constituents, Charity Navigator users, to survey their reactions to the various formats we might use to present this information and their views on the value of this information.

The second part of the answer is a sufficient goal to command our attention – to drive more money to high performing charities and motivate other charities to become high performing. We are not relying on our efforts alone to realize this goal. The results reporting logic model hypothesizes that since by definition the CN 3.0 four-star charities will be transparently reporting their results, other charities will invest in measurement. This will create a virtuous cycle. As more charities are

publicly analyzed and rated, more funding will become available for organizational results reporting and third-party reviews, which will further enhance organizational improvement and donor reallocation toward higher performing nonprofits. The overall outcome will be a significant acceleration in human welfare gains and solving our most grave societal and global challenges. This is one of the ultimate outcomes that Charity Navigator aspires to help achieve.

Glossary

Charity Navigator uses a simple cause logic model to define the common terms used to describe results for charities.

1. **Inputs** are what go in, such as funds, equipment, and human effort and expertise.
2. **Activities** are what gets done, for example providing food, shelter, counselling or training. It could also be the construction of buildings or other material things. It might also be research, policy dialogue or workshops.
3. **Outputs** are the immediate results of the activities – people fed or sheltered, patients treated, trained or educated. Roads are built. Or papers are published and coalitions or networks are built.
4. **Outcomes** refer to what is different as a result of inputs and activities. They may be apparent in the short, medium or long term. Typical outcomes are things like improvements in nutritional status, material welfare, living conditions, and health indicators. They are often measured by changes in income. With respect to policy, they are seen in increased political voice and changes in policies or laws.
5. **Impacts** are the outcomes of an intervention that have been demonstrated to be caused exclusively from an organization’s intervention. Establishing this level of proof of causation is complicated and expensive to do, and of little practical value for most charities most of the time. For that reason, the focus of CN 3.0 is on outcomes.