Social Value Assessment Tool for Nonprofit Organizations In the Social Sector

For use by External Evaluators

7th Revision
Revised June 20, 2009

David E. K. Hunter, Ph.D.
HUNTER CONSULTING, LLC
Hamden, CT USA

Steve Butz
Social Solutions, Inc.
Baltimore, MD USA

Introduction

This tool is designed to enable external evaluators ***trained in its use*** – with explicit attention paid to achieving and assuring inter-rater reliability – to assess the risk that a ***nonprofit social or human services agency***[1] is not able to deliver measurable social value[2] and hence is a very risky organization in which to invest if one cares about such things. The information developed with this tool is intended to be a proxy to help those social investors who want to see measurable good result from their allocation of funds to such organizations. It concretizes the approach to understanding the ability of organizations to create and deliver social value reliably as described in our 2008 white paper *Guide to Effective Social Investing.*[3]

In that paper we identified three performance domains – tactical data use, strategic data use, and program value – that together define the social value of an agency's work. We further identified two indicators for each domain. These are:

1. **Tactical Data Use Domain**
    a. Data Integrity Indicator
    b. Outcomes Focus Indicator

2. **Strategic Data Use Domain**
    a. Making Essential Adjustments Indicator
    b. Relating Staff Efforts to Outcomes Indicator

3. **Program Value Domain**
    a. Capacity to Deliver Program/Services with Fidelity Indicator
    b. Program Impact Data Indicator

We believe that the six indicators that define the three performance domains constitute those aspects of an organization that should be looked at closely when a social investor is attempting to determine the likelihood that an organization is able to generate social value reliably and sustainably.

To assess this capacity to deliver social value we have developed a short but comprehensive set of 24 simple and straightforward questions that, when answered and scored, produce ratings of organizations in each of the three performance domains. ***Clearly, organizations that rate high (in the red) in one or more domains are the riskiest investments.*** That does not necessarily mean they should be avoided. But it does mean that a social investor should proceed with open eyes and the understanding that simply giving such organizations money most likely will not generate social value. And we propose that the intended results of any such risky investment be the development of the organization itself, not (in the foreseeable future) any outcomes it seeks for the people it serves. We call these ***social venture investments.***

---

[1] Organizations working in different domains would need to have a tool with significantly different questions.
[2] An agency's social value consists of the measurable improvement(s) in some aspect of the world that an organization produces. It is not the work that an organization does – it is the *results* of that work.
[3] A copy in pdf format can be downloaded from the website of the *Working Group for Effective Social Investment* at: http://www.alleffective.org/.

Such organizations will need considerable time and significant nonfinancial support – and a high degree of involvement from investors – to help them improve to the point where they are reliable sources of social value. These kinds of high risk social investments (with great opportunity costs) can be considered the equivalent to venture capital investments in start-ups. Examples include many beloved grass roots, community-based, neighborhood, and Faith-based organizations.

Investments in organizations with domain ratings mostly in the medium range (yellow) are somewhat less risky, but most will require no less by way of vigilance, nonfinancial support and investor involvement if they are to mature into high social value institutions. Again, the opportunity costs are significant and should not be underestimated. But, then too, they offer investors the chance to roll up their sleeves and help nonprofits that they find attractive do what is necessary to become high performing organizations – a real contribution to society when done successfully.

By now it will come as no surprise that ***we see only those organizations in the high range (green) in all three domains as "blue chip" social investment opportunities*** – where investors' dollars are essentially purchasing proven social value and there is little need for further investor efforts. However, even with such highly rated organizations there is a need for periodic reassessment. We believe that by applying this method initially, and at least annually thereafter, investors can usefully track the investment types ("blue chip," "social venture") and risk levels across their social investment portfolios, plan the kind and degree of their involvement with the nonprofits in which they invest, and calculate how best to make future investments align with and enhance the social value that they care about.

But to be clear: ***An organization's rating should be seen as something that captures only a snapshot of the organization at a single point in time.*** Organizations that are defined as "risky" investments using the rating tool on a given date may, in fact, be on an upward trajectory over a two or three year period that a social venture investor might reasonably expect to see realized. These organizations are perhaps the best long run "bets" of the social venture investor, provided they continue to improve and when appropriate undertake rigorous program evaluations. Of course, in the same line of thinking, a "risky" organization trending the wrong way through periodic use of the assessment tool should be avoided altogether – unless the investment is made very intentionally to arrest or reverse the organization's decline (in which case there should be a time limit after which no further investments are made if no improvement is evident).

While we have designed this tool for use by external evaluators, in our view it also can and indeed should be used by the leaders of nonprofit social service agencies to assess their own organizations. Thoughtful, honest, judicious answers to the 24 question will result in a clear understanding of an agency's strengths and weaknesses with regard to its social value – and also provide its leaders with clearly identified targets for organizational development.

It is in that spirit that we offer this tool for public use...while hoping that at the same time social investors will look for external evaluators to use it to map the field of social service agencies.

Finally, we do not regard this tool as a finished product. While we have tested many of its questions with several hundred nonprofit agencies, it is very much a work in progress. And we hope that others will use it as a springboard for refining their thoughts about how best to assess the ability of social service agencies to deliver solid social value reliably, measurably, and sustainably.

David Hunter
Steve Butz

May 2009

Nonprofit Social Value Assessment Tool
For use by an external evaluator
Used with the permission of the authors

Date: ______________

Nonprofit organization being assessed: ____

Location: ____

Evaluator / Analyst: ____

Instructions: The external evaluator will base his or her ratings on data collected during a site visit to the organization being assessed. At a minimum, the ratings should be based on:

- a review of the organization's internal performance management (including internal measurement and evaluation) data
- a review of all external evaluation reports
- discussions with the executive director, chief operating officer or director of programs
- discussions with all program directors
- discussions with the development director (and/or anybody who is responsible for reporting data to funders and other external parties)
- discussions with front line staff – at least one person from each program

Summary of Assessment Findings[4]

Summative Domain Ratings	Investment Risk Level Low, Medium, High
Domain I: Tactical Data Use	
Domain II: Strategic Data use	
Domain III: Program Value	
Overall Organizational Rating	

Rationale for Overall Investment Risk Rating[5]

[4] See pages 22 – 28.

[5] Beyond that, we would suggest the following standard: A "low investment risk" rating can only be achieved if an organization has received a "low "rating in all 3 domains; a "high investment risk" rating is assigned if any of these domains has received a "high" rating. All other combinations of domain ratings should result in a "medium investment risk" rating.

Section I
Score all Items[6]

1. The organization measures and analyzes the outcomes for clients produced by staff members' activities/services (including negative outcomes when these become apparent).

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following data were reviewed to make this assessment:

2. Based on available performance data, the organization can be confident predicting service utilization quarter by quarter.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following data were reviewed to make this assessment:

[6] Please Note: Scales vary in the direction of scores based on Items' contents.

3. In the past year, the organization has made a significant change (restructuring, adding/eliminating program(s), program redesign, change(s) to service frequency/intensity) due to knowledge it gained from positive and negative performance data it collected and analyzed.[7]

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

4. Some of the data the organization collects are never or rarely reported on, and may be unnecessary, because they are not related to its mission.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
4	3	2	1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

[7] Changes made for other reasons are not pertinent to this assessment.

5. There are things that the organization believes it is very good at, which it never collects data about or reports on (and therefore can't document).

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
4	3	2	1

The following interviews and associated findings were used to make this assessment:

6. When it is time to generate annual reports or requests for information from funders or other key constituencies, the organization spends what it considers a great amount of time[8] just getting the data into a usable format.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
4	3	2	1

The following interviews and associated findings were used to make this assessment:

Indicate estimates by the people interviewed regarding how much time the organization requires to produce such reports and responds to such requests.

[8] This question is deliberately phrased in a subjective way – it rests on the assumption that organizations which feel burdened by having to spend what they consider to be a great amount of time to producing data outputs are not likely to have a system for generating data that is useful for day to day performance management.

7. The organization is nimble and flexible in pursuing revenues – it will apply for and accept grants, contracts and other funds regardless of whether the activities they support are well aligned with its mission and competencies – and if necessary will hire new staff and build new competencies to do so.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
4	3	2	1

The following interviews and associated findings were used to make this assessment:

8. The organization is highly strategic in pursuing revenues – it will not apply for, and will decline grants, contracts and other funds that in any way involve activities not aligned with its mission and competencies.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used to make this assessment:

9. Growth priority: the organization uses most of its discretionary resources to grow in scale – intending to grow its' program(s)' capacity by multiples over the next 1-3 years (through replication, expansion of current program size, etc.).

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
4	3	2	1

The following interviews and associated findings were used to make this assessment:

10. The organization uses most of its discretionary resources to maintain the quality and fidelity of our program(s) – e.g., with growth a secondary objective.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used to make this assessment:

11. The organization's performance management database records and tracks (check all that apply:
    a. client demographics and baseline data ___
    b. service utilization ___
    c. outcome data ___
    d. staff activities ___

4 items checked	3 items checked	2 items checked somewhat	0 or 1 item checked
1	3	3	4

The following data were reviewed to make this assessment:

12. The organization’s front line staff are held accountable for clients’ progress toward achieving targeted outcomes, and for negative outcomes too.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used to make this assessment:

13. The organization’s front-line staff and their supervisors routinely use individual client data and aggregated caseload data to make decisions and manage performance to ensure fidelity to the model program that has been replicated/implemented.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	2	4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

14. The organization systematically tracks client (program participant) outcomes (positive and negative) until it knows whether participants have achieved what it considers to be long term outcomes (not just short term and intermediate outcomes).

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following data were reviewed to make this assessment:

Indicate the long term outcomes that are tracked, the indicators that are used, and the time period for which the organization tracks such data after participants have completed its program(s).[9]

[9] This also is a way to look at how deeply the organization understands basic performance management concepts – such as, in this case, long term out comes.

15. Staff at all levels of the organization are held accountable for timely and accurate data entry.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

16. The organization routinely conducts data integrity audits and mandates corrective actions as indicated.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

17. Supervision of front-line staff routinely involves routine reviews of the outcomes achieved by the clients they are serving.

Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

Section II

Give the rating that most closely describes the organization

18. The likely positive[10] impact of the organization's program(s) has been:

Proven in at least one randomized control group study by an external evaluator at a 90-95% degree of confidence.	Demonstrated through internal performance data (outcomes) that have been analyzed by external evaluator(s) and compared to outcomes obtained for carefully matched groups of non-participants with at least a 90-95% degree of confidence.	Demonstrated through internal program data that have been benchmarked against similar outcomes for the general population (or key groups) through the use of public data sets.	Suggested by comparing program participant outcomes to duration and dosage of their program participation.	Suggested by tracking program participant outcomes.	Suggested by collecting anecdotes about program participants – or just assumed without collecting any pertinent information at all.
1 or 2 *use the higher number if findings were less statistically valid*	3 or 4 *use the higher number if findings were less statistically valid*	5 or 6 *use the higher number if findings were less statistically valid*	7	8	10

[10] Negative impact data should, if they point to serious threats to program participants' lives and prospects, immediately result in a rating of "high risk" investment…trumping all other considerations. Where some negligible negative impacts are noted, these should be weighed against whatever positive impacts are established…and might result in a decision to rate the organization as a "medium level risk" investment, with the idea that with some hard work the negative impacts can be eliminated and the positive ones strengthened or amplified.

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

Section III

For each item, choose the answer which best reflects the organization

19. The organization's program(s) are evidence-based

   a. None are
   b. One out of several is
   c. Several but not all are
   d. All are

   Score as follows: a = 4, b = 3, c = 2, d = 1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

20. The organization has a robust theory of change featuring:

a. A well delimited domain for action (e.g., a target population)
b. A codified program (with clear elements/activities, personnel competencies, implementation standards, performance standards
c. Measurable outcomes for programs intended to produce outcomes

Score as follows: no elements are present = 4, one element present = 3, two elements present = 2, three elements present = 2, four elements present = 1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

21. If the organization were asked to identify the percentage of clients with certain key demographics (e.g., gender, race, income level) who were enrolled in its program(s) last year, how quickly could it produce accurate data?

    a. There is no way to know (it has never reported on that or the information is missing/incorrect)
    b. In a week or more
    c. In a few days
    d. In one day or less

    Score as follows: a = 4, b = 3, c = 2, d = 1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

22. If the organization were asked to assemble all of the previously recorded performance information (e.g., case notes, enrollment data, attendance/service utilization information, outcomes) on any given individual it has served in its program(s), it would be:

    a. Impossible – The organization doesn't know where all the information is, and some of it may have been lost or misplaced
    b. Difficult – The information is in multiple locations and formats
    c. Not Very Difficult – The information is all in the same place, but is in hard copy form only
    d. Easy – the information is all in the same place and is in digital form

    Score as follows: a = 4, b = 3, c = 2, d = 1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

23. The performance data routinely available to the organization can best be characterized as:

a. Possibly inaccurate, but the organization cannot be sure because it does not monitor data quality
b. The data are highly suspect, and the organization could not feel confident relying on it
c. The data are somewhat accurate, but this varies considerably from quarter to quarter or across programs
d. The data are useful and the organization can usually generate accurate reports

Score as follows: a = 4, b = 3, c = 2, d = 1

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

24. How detailed are the participation/service utilization and progress reports for each program participant?

    a. The organization does not record this type of data on program participants
    b. The organization reports on enrollment
    c. The organization reports on enrollment and attendance at each event
    d. The organization reports on enrollment/attendance and time spent with each program participant
    e. The organization reports on enrollment, attendance, and personal changes (improved attitude, reduced risk of violence, etc.) for each program participant.

    Score as follows: a = 1, b = 2, c = 3, d = 4, e = 5

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

25. The organization measures client outcomes…

a. Never or inconsistently
b. At the end of a program (or when clients' exit from a program)
c. At the beginning of a program **(**or when clients enroll in a program – i.e., at baseline) and at the end of the program (or when clients' exit from program)
d. At specified intervals in addition to the beginning and the end
e. At each client contact

Score as follows: a = 1, b = 2, c = 3, d = 4, e = 5

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

26. To the extent that the organization does measure client outcomes, how easily can it aggregate the data?

a. The organization doesn't know (it does not report on that, or the information is missing/incorrect)
b. In a week or more
c. In a few days
d. In one day or less

Score as follows: a = 1, b = 2, c = 3, d = 4

The following interviews and associated findings were used, and data were reviewed, to make this assessment:

Indicator Scoring Tool

Scoring System Developed by
David E. K. Hunter, Ph.D.
Hunter Consulting LLC

Domain I: Tactical Data Use

Indicator 1: Data Integrity

Pertinent Assessment Items	Evaluator's Assessment Score
#5	
#6	
#15	
#16	
#19	
#20	
#21	
#22	
#23	
Total Score	

Indicator 1: Data Integrity – Total Score Range	**Indicator 1: Numeric risk rating**
9 - 10	**1**
11 - 12	**2**
13 - 15	**3**
16 - 19	**4**
20 - 23	**5**
24 - 27	**6**
28 - 30	**7**
31 - 33	**8**
34 - 36	**9**
37 - 38	**10**

Indicator 2: Outcomes Focus

Pertinent Assessment Items	Evaluator's Assessment Score
#1	
#7	
#8	
#9	
#10	
#11	
#12	
#14	
#17	
#20	
#23	
#24	
Total Score	

Indicator 2: Data Integrity – Total Score Range	**Indicator 2: Numeric risk rating**
12 - 15	**1**
16 - 19	**2**
20 - 23	**3**
24 - 27	**4**
28 - 31	**5**
32 - 35	**6**
36 - 39	**7**
40 - 43	**8**
44 - 46	**9**
47 - 49	**10**

Domain II: Strategic Data Use

Indicator 3: Making Adjustments

Pertinent Assessment Items	Evaluator's Assessment Score
#3	
#7	
#8	
#9	
#10	
#13	
Total Score	

Indicator 3: Making Adjustments – Total Score Range	**Indicator 3: Numeric risk rating**
6 - 7	**1**
8 – 9	**2**
10 - 11	**3**
12 - 13	**4**
14 - 15	**5**
16 - 17	**6**
18 - 19	**7**
20 - 21	**8**
22 - 23	**9**
24	**10**

Indicator 4: Relating Staff Efforts to Outcomes

Pertinent Assessment Items	Evaluator's Assessment Score
#1	
#12	
#17	
Total Score	

Indicator 4: Relating Staff Efforts to Outcomes – Total Score Range	**Indicator 4: Numeric risk rating**
3	**1**
4	**2**
5	**3**
6	**4**
7	**5**
8	**6**
9	**7**
10	**8**
11	**9**
12	**10**

Domain III: Program Value

Indicator 5: Capacity to Deliver Program/Services with Fidelity

Pertinent Assessment Items	Evaluator's Assessment Score
#1	
#3	
#5	
#8	
#9	
#10	
#11	
#12	
#13	
#17	
#20	
#21	
#22	
Total Score	

Indicator 5: Capacity to Deliver Program/Services with Fidelity – Total Score Range	**Indicator 5: Numeric risk rating**
19 - 21	**1**
22 - 24	**2**
25 - 28	**3**
29 - 33	**4**
34 - 37	**5**
38 - 42	**6**
43 - 46	**7**
47 - 49	**8**
50 - 51	**9**
53 - 52	**10**

Indicator 6: Program Impact Data

Pertinent Assessment Items	Evaluator's Assessment Score
#18	
#19	
#20	
Total Score	

Indicator 6: Program impact data – Total Score Range	**Indicator 6: Numeric risk rating**
3	**1**
4	**2**
5	**3**
6	**4**
7	**5**
8 – 10	**6**
11 - 13	**7**
14 - 15	**8**
16 - 17	**9**
18	**10**

ASSESSMENT RESULTS

Indicator	**Level of Investment Risk for Indicator Ratings:** **LOW = rating of 1 - 3** **MEDIUM = rating of 4 – 7** **HIGH = rating 8 - 10**
Domain I: Tactical Data Use	Domain Summative Risk Rating (low, medium, high):[11]
Indicator 1: Data Integrity	Indicator numeric risk rating:
Indicator 2: Outcomes Focus	Indicator numeric risk rating:
Domain II: Strategic Data Use	Domain Summative Risk Rating (low, medium, high)
Indicator 3: Making Essential Adjustments	Indicator numeric risk rating:
Indicator 4: Relating Staff Efforts to Outcomes	Indicator numeric risk rating::
Domain III: Program Value	Domain Summative risk rating (low, medium, high):
Indicator 5: Capacity to Deliver Program/ Services with Fidelity	Indicator numeric risk rating:
Indicator 6: Program Impact Data	Indicator numeric risk rating:

[11] Use the two Indicator Ratings of each domain to calculate a Domain Summative Risk Rating as follows: the combination of 2 Indicator scores in the low range = a Domain Summative Risk Rating of LOW; and the combination of 2 Indicator scores in the high range = a Domain Summative Risk Rating of HIGH; any other combination of 2 Indicator scores = a Domain Summative Risk Rating of MEDIUM.