

IDinsight

# Our first decade of impact



**Internal impact measurement methodology**

September 2022

# Internal impact measurement methodology

## Summary

**Background:** IDinsight was established a decade ago to help non-governmental organizations, government actors, philanthropies, international agencies, and other social sector actors use data and evidence to yield more significant benefits for marginalized communities in Africa and Asia. Our primary goal remains to help our partners maximize their impact. Over the last ten years, IDinsight has grown to 240+ people from 25+ countries, who have worked with 60 non-profits, 30 government partners, 27 foundations, and 38 other organizations.<sup>1</sup> What lessons for the future lie behind these numbers and experiences? What hypotheses might we generate using this data? Which types of projects seemed to generate the most impact? What conditions correlated with success or struggle?

Between October 2021 and April 2022, IDinsight conducted an internal exercise to approximate our projects' contribution to social impact to date (through October 2021). IDinsighters from all regions contributed crucial data, which were then analyzed and reviewed. The purpose of this exercise was to incorporate insights from the data into internal decision-making to increase our future impact, in line with our five-year strategy.

This is just the first step in our evolving internal impact measurement work. In the future, we will continue to improve this methodology and look for more ways across our theory of change where we can generate, analyze, and use such data to improve our decision-making, and thus, our impact.

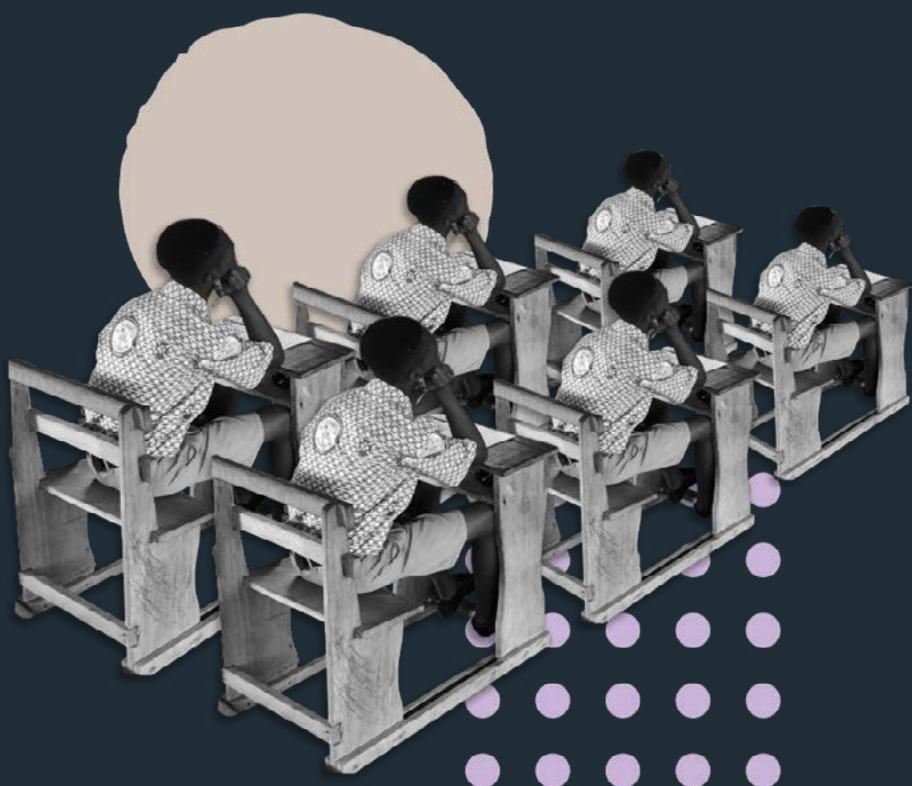
We'd like to thank the team at the Global Innovation Fund, whose framework and materials we heavily borrowed and adapted for this exercise.<sup>2</sup>

1 As of March 2022

2 See their work [here](#)

**Methodology overview:** This internal exercise uses self-reported interviews to estimate IDinsight's contribution to our partners' impact through four distinct pathways, which we call "impact contribution pillars": **Improving Lives, Allocating Resources, Improving Capability, and Advancing Ideas** – of which the largest and most important pillar is **Improving Lives**. To measure the number of lives we contributed to improving, we estimate the breadth and depth of our partners' programs and our contribution to the impact of those programs, supported by different types of evidence. While not perfect, we believe this methodology is right-sized, reflecting our understanding of our pathway to impact, and balancing tradeoffs between rigor, decision-relevance, feasibility, time, and cost.

**Interpreting and using results:** We believe this measurement approach produces results that are at least directionally consistent with IDinsight's impact. The data will be used along with existing information (including that from other quantitative and qualitative sources, as well as expertise and experience) to aid decision-making. The results can help IDinsight teams generate and update hypotheses on trends within IDinsight. They can be referenced when making decisions while developing and executing projects.



# Measurement Framework

## Methodology Overview

Our primary research question is: **What has IDinsight's contribution been to our partners' social impact and how has that varied across projects?** The answers should help us understand when our work makes the greatest difference and also put us in a better position to allocate our own time, funding, and other resources to achieve our ambitious mission.

When approaching this exercise, we attempt to balance the following considerations:

- 1 Rigor & feasibility:** Absent any constraints, we would ideally conduct an experimental evaluation to answer questions about our impact. However, our typical impact evaluation approach is not realistic for this exercise given the context in which we work.<sup>3</sup> Instead, we measure outcomes (and sometimes proxies for those outcomes) along our theory of change that we believe will provide us with a reasonable approximation of our contribution.
- 2 Decision-relevance:** Our approach when working with partners is to produce evidence that informs their decision-making. And for us, insights from this exercise should inform decisions we make internally, like how we can better structure engagements, or how we should evolve our service offerings.
- 3 Time & cost:** To continually inform internal decisions and learning, our approach needs to be relatively quick and affordable, yet still provide meaningful insights. Therefore new data should be relatively easy to collect and frequently updated, similar to data that feeds into a monitoring system.

Through our methodology, each project is ultimately assigned an impact contribution score of zero/negligible, low, medium and high. The figure below illustrates the full process.

**Figure: IDinsight's Measurement Framework**



3 When we conduct impact evaluations, we estimate the impact of typically standardized interventions at a sufficient scale such that we can isolate the estimated impact from noise. We also use a comparison group to isolate the estimated impact of this intervention from other factors affecting outcomes. IDinsight does not currently provide a standardized service/intervention, nor does it operate at a scale sufficient to parse out noise; IDinsight's service does not have an easy comparison group either – we do not know what would have happened had we **not** worked with a certain partner.

# Impact Contribution Pillar(s)

To estimate our contribution, we first identify four main impact contribution pillars:

- 1 **Improving Lives** where we contributed to a partner's direct work to improve lives.
- 2 **Allocating Resources** where we helped a partner improve how they use funds for higher-impact work.
- 3 **Increasing Capability** where we helped improve a partner's ability to produce, use, or advocate for data and evidence in their work.<sup>4</sup>
- 4 **Advancing Ideas** where we directly advanced ideas that were innovative and/or had the groundbreaking potential to improve lives.

Each of the pillars measures total contribution to impact through some combination of the components: [Partner program/policy's] "breadth" and "depth," and [IDinsight's] "contribution". For each project, we measure the impact contribution score for all relevant pillars. A project's ultimate impact contribution score is the maximum score calculated across pillars. The **Improving Lives pillar** is the most direct, the most important, and typically the highest scoring pillar of IDinsight's contribution to impact. This is also the most direct method for measuring total contribution to total impact. We further describe the methodology in the next section.

## Improving Lives

### Components

**Figure: IDinsight's contribution to social impact for the Improving Lives pillar**



For the **Improving Lives pillar**, as shown in the figure above, breadth and depth (i.e. "total impact" of our partners' programs/policies) reflect the number of people impacted by our partners' programs/policies and the magnitude of that impact per person, respectively. It is important to note that breadth is the number of people **actually affected** by the policy, not just those reached. For example, if a government program reaches one million students, but only ten thousand actually learn more as a consequence, the correct breadth is ten thousand, not one million. The corresponding depth will reflect the average learning gains of those ten thousand students.<sup>5</sup>

4 This only refers to projects in which the aim of the engagement was to increase capability. It does not include any increases in capability that resulted indirectly from other collaboration with partners.

5 This is similar to the concept of the "Treatment on the Treated (ToT)" estimate, vs the "Intention to Treat (ITT)" estimate. We use ToT for both breadth and depth.

Contribution reflects how much of that total impact is due to decisions made, and actions taken, based on the data, evidence, systems, tools, and/or recommendations ("deliverables") produced by IDinsight (moderated by the level of our own confidence in the deliverables).

## Component Values

Each of the components is measured using either absolute numbers reflecting true values, or categorical values (similar to Likert scales), which are proxies for true values.

- Breadth is simply measured as the number of lives impacted by our partner's programs/policies.
- Depth is measured on a 0-4 scale.
- Contribution is measured on a 0-3 scale.

The table below describes the depth scale in more detail. The following table describes the contribution scale in more detail.

**Table: Depth Component Values for the Improving Lives Pillar**

Depth Value	Rubric			
	% increase in annual income		Health	Education
	one-time	permanent		
0	<50%	<5%	no/negligible improvement	<0.5 years
1	50-199%	10%	minor improvement	0.5-2 years
2	200-700%	50%	major improvement	2-7 years
3	>700%	100%		7+ years
4			life-saving	

**Table: Contribution Component Values for the Improving Lives Pillar**

Contribution	Rubric
0	Partner's decisions/actions are not based on our advice/data. Or, even if their decisions/actions were based on our advice/data, we have very little confidence in our recommendations or support (e.g. they are based entirely on intuition as opposed to data).
1	Partner's decisions/actions are partially based on our advice/data. Or, even if decisions were largely based on our advice/data, limitations in the analysis did not permit us to provide recommendations or support that are strongly evidence-based (i.e., recommendations informed by evidence, but the evidence is limited or unclear).
2	Partner's decisions/actions are largely based on our advice/data and the analysis we did has permitted us to provide recommendations or support that are strongly evidence-based.
3	Partner's decisions/actions are entirely based on our advice/data and the analysis we have been able to do has permitted us to provide recommendations or support that are strongly evidence-based.

## Standardized Impact Metrics: Relative Impact Unit (RIU)

For in-depth analysis for our own internal decision-making, we chose to create a synthetic measure to compare low/medium/high impact projects across our portfolio. While these are not presented on the public microsite because they do not permit intuitive interpretation, we are sharing the methodology for creating standardized metrics below for those who may be interested.

Before combining the individual component values (of breadth, depth, and contribution) to produce an impact contribution score, we apply “moral weights”<sup>6</sup> to scale the relative significance of each component, and each numerical value within depth and contribution. To determine moral weights, we used benchmarks where available.<sup>7</sup> Where benchmarks were unavailable, we ran a sensitivity analysis with different possible values. We call these weights component-level “**Relative Impact Unit**” (RIU)<sup>8</sup>

- RIU<sub>B</sub>: RIU of Breadth<sup>9</sup>
- RIU<sub>D</sub>: RIU of Depth<sup>10</sup>
- RIU<sub>C</sub>: RIU of Contribution<sup>11</sup>

For the Improving Lives pillar, RIU of Breadth, RIU<sub>B</sub> equals the Breadth value itself—the number of people impacted. And the RIU for contribution, RIU<sub>C</sub>, equals the contribution value as well—taking on values 0-3. Each Depth value has its own RIU conversion, shown in the table below.

**Table: RIU values for Contribution for the Improving Lives Pillar**

Depth Value	Weight
0	0
1	286
2	1,430
3	2,860
4	14300



<sup>6</sup> We borrowed this idea from GiveWell.

<sup>7</sup> Global Innovation Fund, GiveWell and IDinsight's Peoples' Preference Project, which was funded to input into GiveWell's calculations.

<sup>8</sup> Modelled closely after Global Innovation Fund's Practical Impact Units. Note that the RIU conversions are different for different impact contribution pillars.

<sup>9</sup> For the Improving Lives pillar, RIU of Breadth = Breadth (i.e. the number of people impacted).

<sup>10</sup> For the Improving Lives pillar, Each Depth value has its own RIU: 0 → 286; 1 → 1,430; 2 → 2,860; 3 → 14,300.

<sup>11</sup> For the Improving Lives pillar, Each Contribution value = the RIU: 0 → 0; 1 → 1, etc.

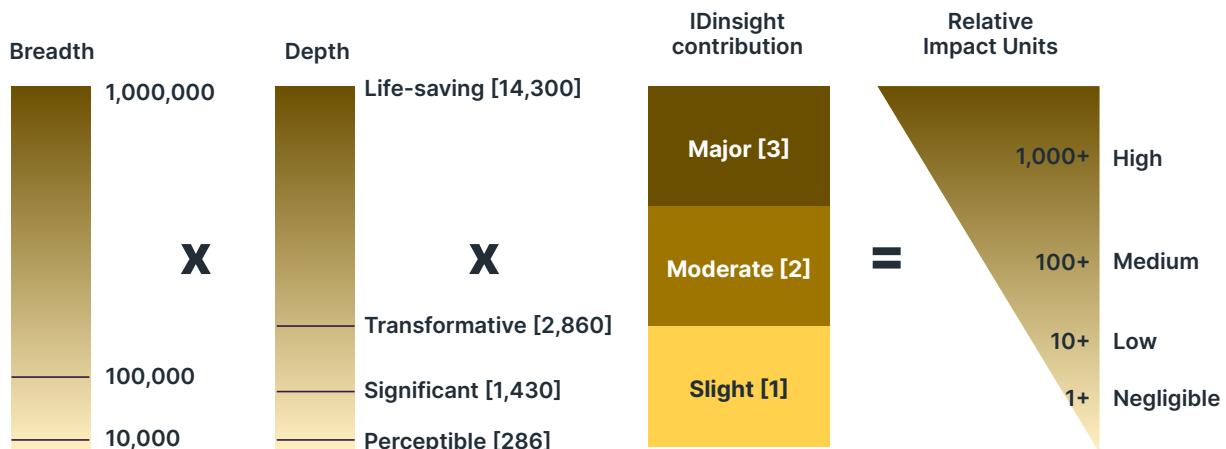
## Total RIU and Impact Contribution Score

The final total metric,  $RIU_T$  for each project is a function of all component-level RIUs. We divide the product of the component-level RIUs by 100,000 to make the resulting numbers easier to compare.

$$RIU_T = RIU_B \times RIU_D \times RIU_C \div 100,000$$

Once we have the final values RIUT, we convert those back into categorical values of impact contribution: zero/negligible, low, medium and high. The entire process is described in the figure below.

**Figure: Relative Impact Units (RIU), [Values], and Final Impact Contribution Score<sup>12</sup>**



## Allocating Resources, Increasing Capabilities and Advancing Ideas

The approach used to measure contribution to impact in the Improving Lives Pillar is the most direct. For the remaining pillars—Allocating Resources, Increasing Capability, and Advancing ideas—two or more of the “depth, breadth, and contribution” components are converted into similar scores appropriate for the respective pathways.

We also use different moral weights for each component. We compared the results to intuition to select the final moral weights. Moral weights for Capability Building and Advancing Ideas are extremely rough estimates. No external benchmarks are available. These weights are based on our intuition of the relative impact of projects.

# Data

## Data Collection

For each project, and for each relevant impact pillar, we measure component (Breadth, Depth, Contribution) values based on a combination of self-reported and publicly available evidence. Breadth and depth are estimated using client data, publicly available literature, and/or evaluation results on our partners' programs. Our contribution is subjective self-assessment backed up by informal evidence, such as client conversations, about the degree to which our contributions were useful and used in the pursuit of impact.

Between November 2021 and March 2022, we administered 162 interviews<sup>13</sup> to project leads, capturing their reflections and insights. We collected data for 99% of projects that ended between 2011 and 2021.

## Data Quality

Interview responses went through the following checks:

**Review for evidence strength:** The scores for each project need to meet an evidence threshold to be officially counted towards our impact. Each submission for breadth, depth, and contribution provides evidence supporting the score.<sup>14</sup> We then score this evidence against a rubric. In cases where the evidence is inadequate,<sup>15</sup> we ask the respondent to provide more evidence. If the evidence strength remains inadequate, the calibrated score of the project becomes "Don't know" instead of the initial score given.<sup>16</sup>

**Review for consistency:** We review each response to check whether it is consistent with the intended framing of the question. Where we find issues, we ask for clarifications and make changes. Given the broad range of project types at IDinsight, fitting projects into a narrow set of questions is understandably a challenge for respondents. We sometimes have to make judgment calls about how exactly to classify a project.

13 Not all data was collected as interviews. After a project lead felt comfortable with the questionnaire, they often filled out responses for additional projects themselves.

14 Example sources of evidence included publicly available information, studies done by IDinsight or other organisations, or the client's actions.

15 According to rubrics of evidence strength, evidence strength would be inadequate if the evidence was below weak, did not support the score, or was unclear.

16 Exceptions: If a project's original score was zero, the calibrated score of the project would still be "Zero".

# Using and Interpreting the Results

## Using the Results

Building comparable impact estimates gives us the foundation to run many different types of analyses, sometimes including additional data. We've calculated topline impact estimates by year to help tell the story of our contribution to impact over time. We use project cost-effectiveness as the basis for internal analysis on what's worked, and where we can likely do better.

**Calculating topline impact estimates:** Topline impact estimates show the number of lives and US dollars IDinsight has contributed to improving and allocating. They represent the aggregate breadth estimates where depth and contribution were above a minimum threshold. For Improving Lives, the topline estimate represents the number of lives improved (in a small to life-saving way) in projects where our data/advice at least partially informed our partners' decisions and actions. For Allocating Resources, the topline estimate represents the number of US dollars allocated where our data/advice at least partially informed our partners' decisions and actions.

**Calculating cost-effectiveness:** Impact per cost is our most important metric, as our goal is to make decisions about where we can have the highest relative impact with our time and resources. To determine cost-effectiveness, we divide the projects' impact by the projects' financial cost (RIUT/\$). Cost data is sourced from financial records. While not included in the impact microsite, we use this for our own internal decision-making.



## Interpreting the Results

We acknowledge several limitations in our approach and therefore results. For example:

- **Weak counterfactual:** Through our contribution metric, we implicitly make the assumption of what our client would have done without the evidence IDinsight helped produce. However, this comes with the assumption that had we not taken on the project, no one would have supplied evidence. We do not consider a counterfactual if another organization had done the work instead of us.
- **Self-reported data:** While we tried to reduce biases intrinsic to self-reported data by requiring evidence to support claims made by respondents, the data we used are vulnerable to a number of different challenges.

Given the limitations in our approach, results warrant careful interpretation. We think the approach taken was an appropriate balance of constraints and considerations. We had to make tough trade-offs in building the impact contribution measurement framework, running data collection, and calibrating and analyzing results.

We believe this measurement approach produces results that are directionally consistent with IDinsight's impact. They represent self-reported estimates of our contribution to our partners' impact.

We also acknowledge limitations in how we can interpret the results. Differences in impact measured across regions, sectors, or methods are due to a variety of factors. For example, we cannot simply compare the impact between education and health, and determine that project work in one is generally more cost-effective than the other. That may be a factual historical statement (albeit, with uncertainty given the limitations in our approach), but it is neither a causal statement (working in one sector leads to more impact than working in the other), nor a predictive statement about the future.

# Planned Future Improvements

**Client feedback:** We are building a process to incorporate clients' perceptions of our contributions to their programs into our measurement system. Clients have a stronger understanding of how our contributions were applied, and their opinions will likely be more independent than those of IDinsighters. We will also ask clients what they would have done had the engagement with IDinsight not happened (e.g. make decisions without new evidence or work with another firm that would produce evidence instead of IDinsight).

**Measurement consistency across project lifecycle:** We will collect this data over the life of projects in the future. During the concept stage, we will ask project teams to estimate these parameters. We will then track how they change over the life of the project, and after the project closes. We will assess how accurate initial project estimates are, and what information seems to facilitate good versus poor estimates.

**Data collection at multiple periods after the project closed:** We hope to collect data multiple times after a project closes, perhaps at the time of project close and 12 and 24 months out. Collecting data months or years after projects close will allow time for our contributions to integrate more fully into our partners' work, thus giving us a more accurate picture of our value. We will need to balance the desire for lots of data with the burden this puts on data inputters (team members and clients).



# Appendix

## Appendix A: Calculating RIU for Allocating Resources

**Calculation:  $RIU_T = RIU_B \times RIU_D \times RIU_C \div 100,000$**

- $RIU_B$  = total amount influenced in USD.
- $RIU_C = 1, 2, \text{ or } 3$ , based on the contribution category in the table below.

Table: Contribution weights and rubrics for Allocating Resources pillar

Contribution	Weight ( $RIU_C$ )	Rubric
0	0	Partner's decisions/actions are not based on our advice/data. Or, even if their decisions/actions were based on our advice/data, we have very little confidence in our recommendations or support (e.g. they are based entirely on intuition as opposed to data).
1	1	Partner's decisions/actions are partially based on our advice/data. Or, even if decisions were largely based on our advice/data, limitations in the analysis did not permit us to provide recommendations or support that are strongly evidence-based (i.e., recommendations informed by evidence, but the evidence is limited or unclear).
2	2	Partner's decisions/actions are largely based on our advice/data and the analysis we did has permitted us to provide recommendations or support that are strongly evidence-based.
3	3	Partner's decisions/actions are entirely based on our advice/data and the analysis we have been able to do has permitted us to provide recommendations or support that are strongly evidence-based.

## Appendix B: Calculating RIU for Increasing Capability

**Calculation:  $RIU_T = RIU_{B1} \times RIU_{B2} \times \sqrt{RIU_{B3}} \times RIU_D \times RIU_C \div 100,000$**

- $RIU_{B1}$  = Number of people whose capability increased.
- $RIU_{B2}$  = Seniority of people whose capability was increased. 1-4 scale, [1] for junior staff, [4] for senior leadership.
- $RIU_{B3}$  = Size of population affected by that organization's programs or policies.
- $RIU_D$  = 0, 72, 143, or 286 based on the depth category in the table below.
- $RIU_C$  = 1, 2, or 3, based on the contribution category in the table below.

Table: Depth weights and rubrics for Increasing Capability pillar

Category	Weight ( $RIU_D$ )	Rubric
0	0	No or minimal improvement
1	72	Slight improvement
2	143	Moderate improvement
3	286	Large improvement

Table: Contribution weights and rubrics for Allocating Resources pillar

Category	Weight ( $RIU_C$ )	Rubric
0	0	Partner's increase in capability is not because of your IDinsight's support
1	1	Partner's increase in capability is partially because of IDinsight's support
2	2	Partner's increase in capability is largely because of IDinsight's support
3	3	Partner's increase in capability is entirely because of IDinsight's support

## Appendix C: Calculating RIU for Advancing Ideas

### Calculation: $\text{RIU}_T = \text{RIU}_B \times \text{RIU}_D \times \text{RIU}_C \div 100,000$

- $\text{RIU}_B = 0, 5,000, 20,000, \text{ or } 50,000$  depending on the breadth category in the table below.
- $\text{RIU}_D = 0, 286, 1,430, \text{ or } 2,860$  based on the depth category in the table below.
- $\text{RIU}_C = 1, 2, \text{ or } 3$ , based on the contribution category in the table below.

Table: Breadth weights and rubrics for Advancing Ideas pillar

Category	Weight ( $\text{RIU}_B$ )	Rubric
0	0	The idea was not discussed or applied by any individuals or organizations in the sector.
1	5,000	The idea was lightly discussed by other organisations and actors in the sector (e.g. small twitter buzz/1-5 google scholar citations/1-5 media mentions/50+ backlinks, etc.) AND/OR such discussions lasted for a few months and quickly died down AND/OR the idea was meaningfully put into action (changed beliefs, practices, and programs) by 1-2 organizations.
2	20,000	The idea was meaningfully discussed by other organisations and actors in the sector (e.g. good twitter buzz/6-20 google scholar citations/6-20 media mentions/100+ backlinks) AND/OR such discussions were somewhat sustained and lasted for over a year AND/OR the idea was meaningfully put into action (changed beliefs, practices, and programs) by 3-5 organizations.
3	50,000	The idea was extensively discussed by other organisations and actors in the sector (e.g. Significant discussion on twitter/More than 20 google scholar citations/More than 20 media mentions/200+ backlinks, etc./Included in books, academic curriculum, cited by other organisations as foundational to approach AND/OR such discussions were sustained and persisted for a long period of time (at least 2 or more years) AND/OR the idea was meaningfully put into action (changed beliefs, practices, and programs) by 5+ organizations.

**Table: Depth weights and rubrics for Advancing Ideas pillar**

Category	Weight (RIU <sub>D</sub> )	Rubric
0	0	The idea is not impactful. It has no potential to contribute to improving lives, does not contribute anything new to the discourse, AND/OR is not important in improving the functions of organizations.
1	286	The idea is somewhat impactful. It has the potential to make small improvements to peoples' lives, contributes something slightly new and innovative to the discourse, AND/OR is only somewhat important in improving the functions of organizations.
2	1,430	The idea is moderately impactful. It has the potential to make noticeable improvements to peoples' lives, contributes something new and innovative to the discourse, AND/OR is instrumental in improving the functions of organizations.
3	2,860	The idea is highly impactful. It has the potential to make large improvements to peoples' lives, contributes something ground breaking/highly innovative to the discourse, AND/OR is highly instrumental in improving the functions of organizations.

**Table: Contribution weights and rubrics for Advancing Ideas pillar**

Category	Weight (RIU <sub>C</sub> )	Rubric
0	0	The idea did not originate from your project and/or we had nothing to do with its inception.
1	1	Your project was one of many organisations advancing the idea alongside many other groups/organizations
2	2	Your project, along with 1-2 other groups or organisations, is leading the discourse on the topic. It contributed a significant amount of the relevant evidence and ideas.
3	3	Your team is leading the discourse on the topic. Most or all of the relevant evidence was produced and distributed by your team.

## Appendix D: Example Calculations

The following are example calculations. They all result in the same Relative Impact Units, thus illustrating how comparisons are made across pillars.

**Direct Impact:** IDinsight has a moderate contribution to a program that increases annual income by 50% forever for 500 people.

- Rubrics: breadth [500] \* depth [2] \* contribution [2]
- RIU (using weights): breadth [500] \* depth [1,430] \* contribution [2] / [100,000] = 14 RIU

**Allocating Resources:** IDinsight slightly improves how a foundation improves how they spend \$1,400,000 USD.

- Rubrics: breadth [1,400,000] \* contribution [1]
- RIU (using weights): breadth [1,400,000] \* contribution [1] / [100,000] = 14 RIU.

**Increasing Capability:** IDinsight slightly contributes to a moderate increase in capability of 30 M&E staff that reach 100k people.

- Rubrics: breadth [30 people, their program influences 100k people, junior staff] x depth[2] x contribution [1]
- RIU (using weights): breadth [30 \* sqrt(100,000) \* 1] \* depth [143] \* contribution[1] / [100,000] = 14 RIU

**Advancing Idea:** IDinsight contributes additional evidence about a somewhat influential idea that many other organisations already support. Our publication generates a media mention and is linked to ~75 times.

- Rubrics: breadth [1] x depth [1] x contribution [1]
- RIU (using weights): breadth [5,000] x depth [286] x contribution [1] / [100,000] = 14 RIU

## **Acknowledgements**

We would like to thank the core IDinsight team who worked on this project, Emily Coppel, KJ (Kejian) Zhao, Leonard Oruko, Marc Shotland, Nicholas Parker, Rob Sampson, Ron Abraham, Ruchika Joshi, and Ruth Levine. We would also thank the many current and former IDinsighters who submitted and reviewed data, helped shape our approach, and engaged with us throughout this process.

**IDinsight**

**[www.IDinsight.org](http://www.IDinsight.org)**

**@IDinsight**