The Robustification of Success: How Quantification Transforms Tenuous Impact Claims

Robustification is a process which enhances confidence in evidence by increasing perceived certainty and reliability while reducing critical analysis.

Framework: How to understand the role of numbers in success stories

Through narrative policy analysis I understand the role of success stories as strengthening the evidence-base vis-a-vis competing alternatives by seemingly limiting uncertainty, complexity and ambiguity (Roe 1994).
This way, the case put forth seems less complicated to policy-makers, and thus more doable and less likely to fail.

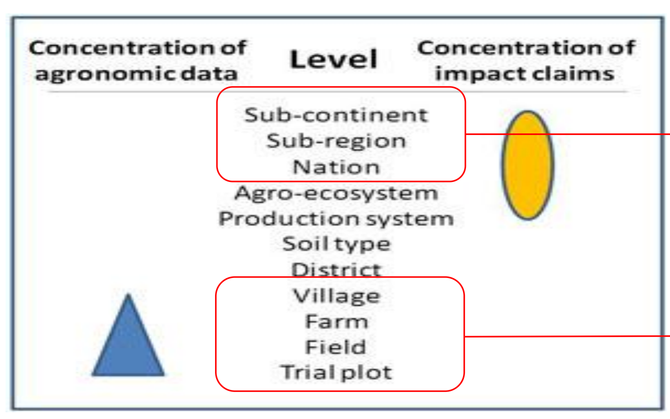
Such "success-making" allows individuals and organisations to justify calls for additional financial resources (Sumberg et al 2012).

Quantification lies at the heart of robustification because it can convey evidence in a manner that increases the chance of it being accepted. Not only does it simplify and depoliticise (Porter 1994), but it induces so-called *peripheral cognition*.

Peripheral cognition implies that our brain is generally programmed to rely on the credibility of the source when presented with numbers due to inadequate numeracy levels (Yalch & Elmore-Yalch 1984). In a way, fact-checking is 'outsourced' to the communicator. Numeracy should not, however, be equated with intelligence, and we are more susceptible to it than most would care to admit (Reyna 2000).

This is problematic because it enhances confidence in an approach by ignoring, instead of constructively engaging with, uncertainties and contestations.

My hypothesis is that quantification, as a tool of success stories, often ends up obscuring the disjuncture between sector-wide claims of impact, and the results from the small-scale agronomic trials on which they are based.



This is the level of claims made in the case studies.

This is the level of research upon which the claims are based.

Source: Levels of Claim Disjuncture, Sumberg 2013

I use pro-poor agriculture as an example because it relies on seemingly simple measurable indicators and outcomes, but which on closer inspection reveal a host of uncertainties, ambiguities and complexities

Agricultural Contentions		
Income	Nutrition	Yields
Complexity ▪ Material/service reciprocity instead of monetary exchange *Uncertainty* ▪ Payment time lag ▪ Volatile prices ▪ Distribution	*Complexity* ▪ Physical measurement ▪ Attribution ▪ Uptake vs retention ▪ Controlled vs uncontrolled environment *Ambiguity* ▪ Caloric sufficiency ▪ Food security ▪ Poverty lines	*Complexity* ▪ Harvest by need vs by season, the former not easily quantified *Uncertainty* ▪ Quality ▪ Distribution ▪ Spoilage in storage ▪ Market access

Sources: Picciotto 2012, White 1999, Masset et al 2011, Jerven 2013, Chambers & Mayoux 2005, Eyben 2013.

This poster will demonstrate how a case study uses numbers to robustify its evidence, as well as point out the weaknesses which might be obscured by it.

Case Study 1: Aquatic Agricultural Systems

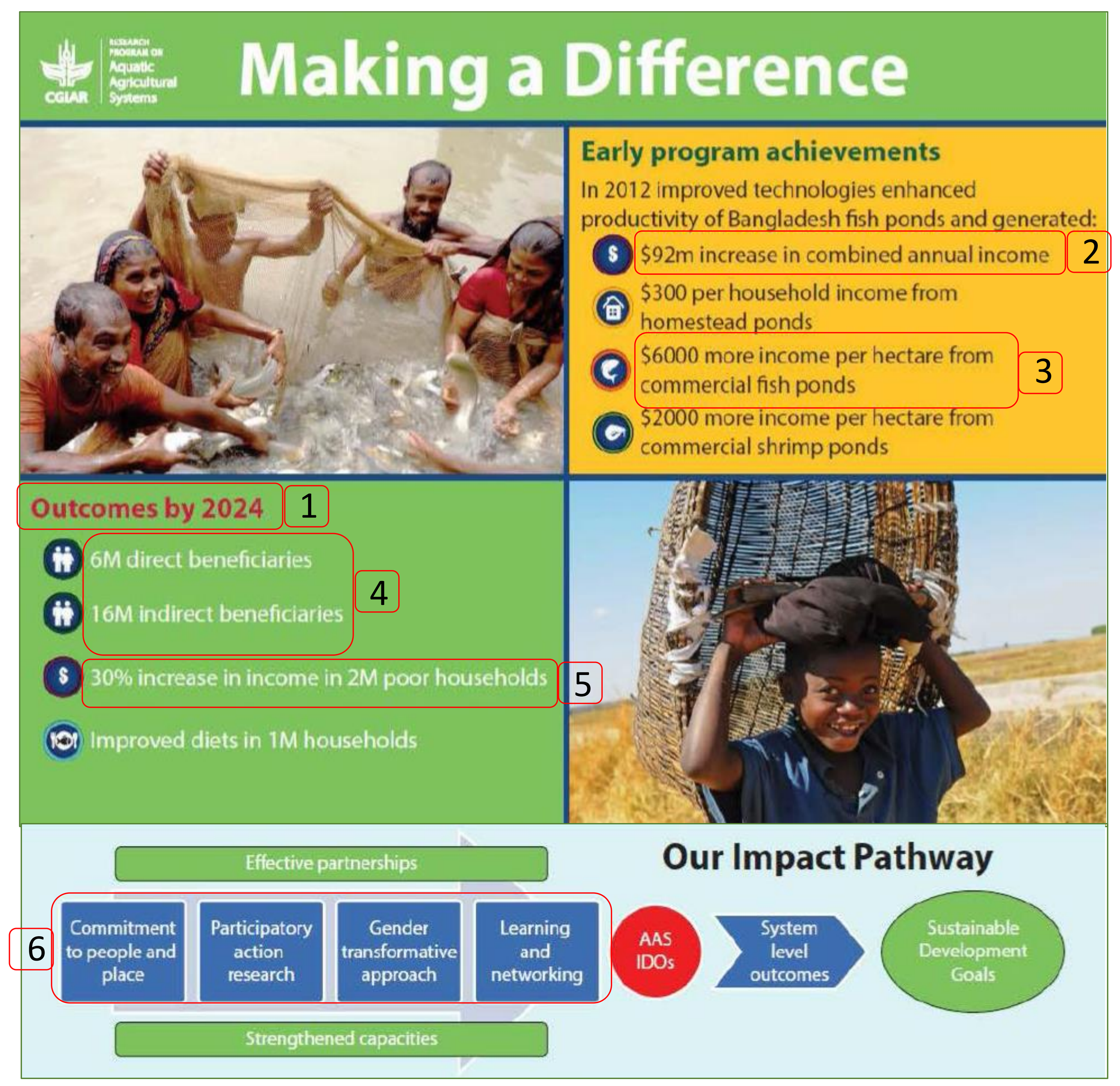
1 The outcomes are in themselves a peculiar feature. Based on income data for one year (2012) they seemingly make claims about impact 12 years into the future. The latter two involve household income and nutrition, which are notoriously hard to measure and both claims are double aggregations.

2 Unless we know how many people this aggregation is based on, and how wealth is distributed, the result remains unclear.

3 Who operates and benefits from commercial efforts. Looking into how income generation is distributed across communities and if any external actors are benefitting or controlling local efforts, would give a more nuanced understanding of their 'success'.

4 Getting big results is easier with a small sample, implying the larger the claim the more tenuous it may be (Huff 1957). Sample size is of utmost importance to correctly appreciate these claims, yet it remains unknown. Considering their beneficiaries range from 6 million (direct) to 16 million (indirect) their sample size is likely much lower and a substantial amount of estimation is involved.

5 "If we take the notion of poverty as a multi-dimensional concept seriously then we should use a range of poverty indicators" (White 1999: 504). Considering the social transformations promised (empowerment, gender equality, and sustainability) AAS ostensibly do. It is therefore odd that their indicators of such change are wholly economic (with the exception of improved diets) and based solely on income. Thus, there is a clear disjuncture between numerical achievements and the social transformations claimed



Case Study 2: AfricaRice

AfricaRice

8

Improving food security

By 2020, sub-Saharan Africa's (SSA's) rice paddy production will have increased from 18.4 million tonnes (Mt) (11.9 Mt of milled rice) in 2010 to 46.8 Mt (30.4 Mt of milled rice), with the productivity-enhancing research and development activities described in this strategy. Without this productivity-enhancing R&D, the baseline levels of paddy production under the 'baseline scenario' (projecting each country's production on the basis of 1980–2010 growth rates) would be 32.3 Mt (21.0 Mt

By improving rice processing technologies and reducing losses, it is expected that the quality of locally produced rice will be increased, generating more revenue for rice processors and rice traders. These benefits are estimated at $64.2 million annually (cumulative 5%-discounted, $323.7 million) for rice processors and $30.8 million annually (cumulative 5%-discounted, $155.3 million) for rice traders.

7

7 Sharp numbers (e.g. 3, 7, 13) implies exact knowledge, gaining a higher level of believability and confidence. Schindler & Yalch (2006) found that stating "our alumni earn between 43,000 and 57,000 a year" was more believable than "they are well-paid", "earning around 50,000", or "earning between 40,000 and 60,000". Similarly, if you were asked to count people in a room and there were 100, you would probably say "exactly 100" as not to be misconstrued as inaccurate (Dehaene 1997).

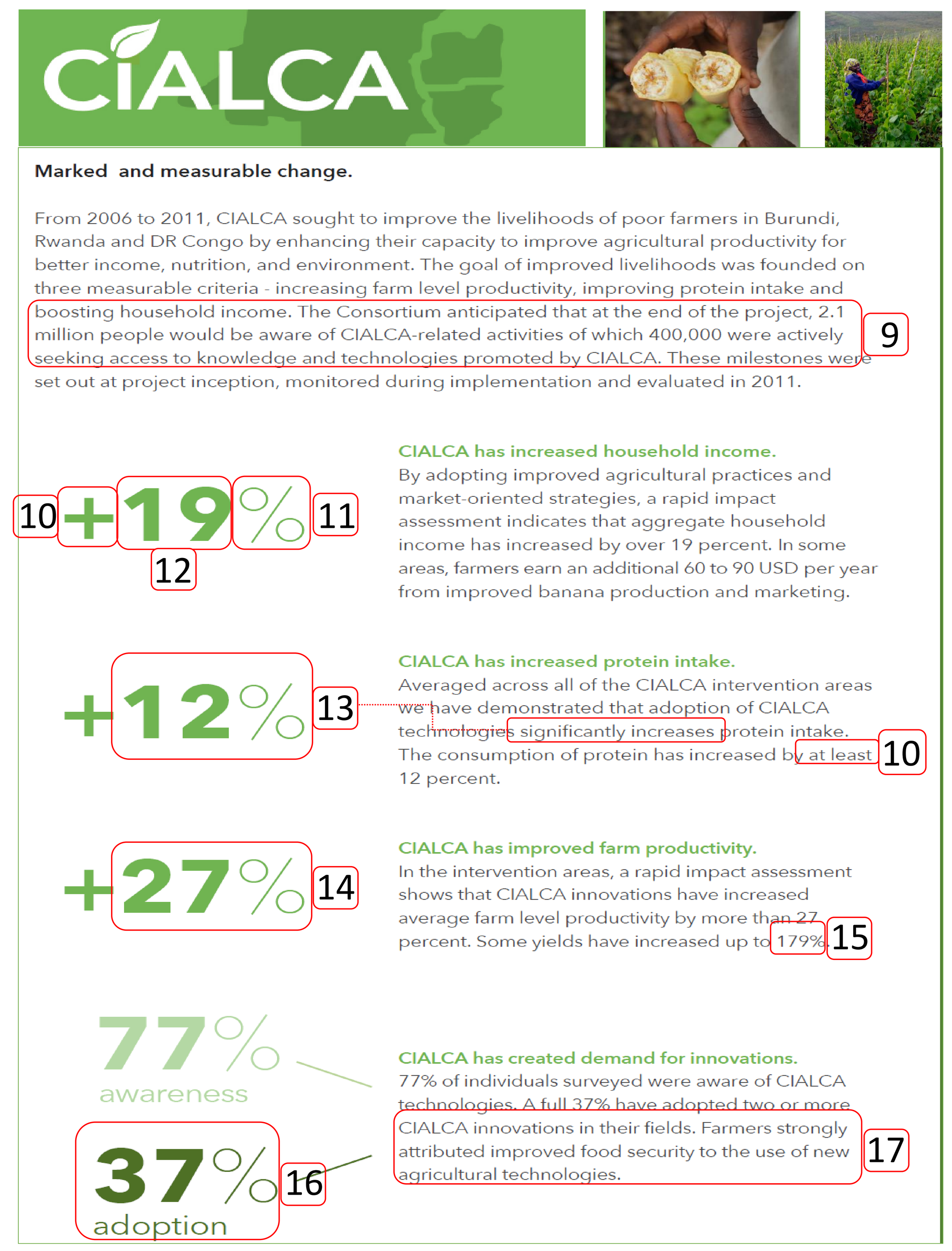
8 The planned intervention will increase yields to 46.8 tonnes. Compared to the 'baseline-scenario', which offers a mere 32.3, a 44.9% increase is promised. Of course 44.9% is better, but this assumes that only one alternative exists and that it is inevitably inferior, effectively closing down discussion.

Case Study 3: CIALCA

9 Absolute numbers (not percentages) are recommended to contrast large numbers. 400,000 direct beneficiaries sound much better than a participation-rate of 19% (of 2.1 million) although communicating the same relationship. Notice the vagueness of terms used, in contrast to the sharpness of results below.

10 Plus-signs precede results, indicating (without mentioning the range) that results are likely to be even greater, displaying a case of 'subtle interpretive work' (Potter et al 1991). The same effect is achieved with 'at least'.

11 Percentages are found to be more 'understandable'. This does not mean percentages make it easier to understand the actual reality behind the statistic, but that whatever information given in percentages will be met with greater acceptance (Brase 2002). They are also easier to remember, and susceptible to deceptive framing (Heath & Heath 2007; Brase 2002).



12 The font used is four times the size of the text, highlighted in colour. Future research could benefit from testing whether such visual presentation facilitates peripheral cognition as it intensifies the reader's focus on numbers.

13 Nutritional stats make for questionable aggregates, and hunger is "averaged across all areas". One could also question if 12% is indeed such a great increase, as 12% of nearly nothing is also nearly nothing. Without a clear baseline the number means little. Furthermore, according to what standard is this result significant?

14 Some food-stuffs like grains can be measured more easily as they are harvested at set times. Cassava, legumes and bananas on the other hand are harvested based on need, i.e. over a season, and is thus not easily quantified (Jerven 2013).

15 An extreme outlier? See 18 for details.

18 Using median (not mean or mode) obscures distribution of land resources, failing to capture high concentrations of wealth. Mean is used when wanting large numbers and median if wanting smaller, "a trick commonly used, sometimes in innocence but often in guilt" (Huff 1954: 30). Targets small-holders, the latter is desirable. It masks incomes from larger units which are generally more attuned to intensification and market approaches, potentially using the enrichment of the well-off to claim improvements for the poor.

16 Framed another way, as their target population is 0.004% of the total (90 million) it turns a 37% adoption-rate into a regional result of 0.0015%. Though an exaggerated example it does demonstrate the power of numerical framing.

Metric		Burundi	DR Congo	Rwanda
Median farm size[3] (ha)	18	0.4 - 1	0.4 - 1	1 - 1.6
food insecurity, 'occasionally' or 'often'[3] (% of households)	19	58	61	40

17 'Strongly attributed' hardly reflects a robust measurement, but is contrasted by a sharp percentage which serves to enhance its perceived level of certainty.

19 'Food-insecurity' is an elusive indicator. Even vaguer is its measurement, being based on the ambiguous terms 'occasionally' and 'often'. Statistics are based on household rather than individual data, showing a double aggregation although the indicator measured is inevitably individual. Quantifying food-security in percentages lends an aura of concreteness to something inherently equivocal.

19 Saying that results 'often' occurred obscures sample size, actual results, and is wholly subjective as there is no universal definition of 'often'.

Implications

Scrutiny reduced: Although audiences are not oblivious to the shortcomings of quantification, its obfuscating tendencies can potentially be ignored as peripheral cognition lowers scrutiny.

Simplification: Through the use of these numerical indicators an overly simplistic notion of development, and what constitutes success, is projected.

Depoliticisation: Numbers have a unique way of making the intangible (e.g. empowerment) tangible (e.g. income), and as the latter is open to empirical testing it gains greater acceptance. Through this process, a higher level of certainty is projected which would otherwise not be achieved. It allows crucial questions to be left unanswered.

Validation: Agriculture may be a field that is easily quantified (yields, income and reach) but that does not mean such quantification always provides meaningful measures or that measures are actually obtainable. The scientific certainty derived from an apparently carefully calculated number might also strengthen the indicator itself, sheltering it from scrutiny.

Standardisation: Following from this we can theorise a process of standardization in which the importance of these indicators are 'normalized'. Internalised by practitioners, communications intermediaries, and beneficiaries, a particular understanding is consolidated, leading to what Porter (1995) calls mechanical objectivity; when practice is consolidated to the point where its underlying assumptions becomes beyond everyday scrutiny.

Self-reinforcement: The result is a self-reinforcing mechanism where the telling of success-stories reifies certain statistics while the use of statistics makes such stories easier to tell.

Further Research

- Does use of colour, font sizes and other graphical techniques facilitate peripheral cognition?
- Beyond peripheral cognition, to what extent is scrutiny reduced by organisational conditions, blind faith in the scientific methods used, time-pressures and mutually self-serving interest between practitioners?
- Time and attention spent on success stories, of both quantified and qualitative formats, and how it affects comprehension and scrutiny.
- Do more credible actors publish more success stories?
- As peripheral cognition depends on credibility, do more credible actors use more quantification?
- An organisational study of the construction, publication, and circulation of success-stories could aid our understanding of how they affect policy.
- To what extent do 'info-mediaries' (Fisher & Vogel 2008) consciously use robustification as a tool?
- Quant speak: To what extent might quantitative jargon induce peripheral cognition?