

Core Methods in Evaluation – Transport Canada

Methodology and Foundations Part 1

Greg Mason
PRA Inc. & University of Manitoba

Ottawa, January 31 – February 1, 2013

Outline

Day 1: (March 31) Evaluation design and planning

Morning - Foundations

- Rationale and scope of government programs/policies
- TB Policy on Evaluation
- Creating an informative logic model and evaluation matrix
- Case example – 2010 Olympics evaluation matrix

Afternoon – Understanding and assessing relevance

- Literature reviews
- Document reviews
- Senior Management and Expert Consultations
- Case example – Labour market information for Aboriginal youth

Day 2: (February 1) Quantitative and qualitative data collection and analysis

Morning – Quantitative methods

- Survey questionnaires (samples, framing questions, modes of data collection, analysis options)
- Basics in measurement – refresher (univariate, bivariate and multivariate)
- Case examples – Public Opinion Research for the 2010 Olympics

Afternoon – Qualitative methods

- Key informant interviews
- Focus groups
- Case studies
- Case example – NCB Focus Groups
- Case example – Media Scan for the 2010 Olympics
- Concluding note - Analysing qualitative/quantitative data – mixed methods

Day 1

Evaluation design and planning

Foundation

Morning
January 31, 2013

Core issues have transformed for federal evaluations

Reduced from

- Rationale/relevance
- Design/delivery
- Success/impacts
- Cost effectiveness/alternatives



To

- Relevance
- Performance

Policy on Evaluation

Goal

1. Defines the obligation for departmental evaluation plans to demonstrate progress toward achieving coverage of direct program spending over five years
2. Plans that do not demonstrate evaluation coverage of all direct program spending need to use a “risk-based” approach to planning coverage

The evaluation plan needs to either show 100% coverage or identify the programs that will be assessed (and not assessed) within the 5-year cycle using a **risk-based criteria**.

Relevance

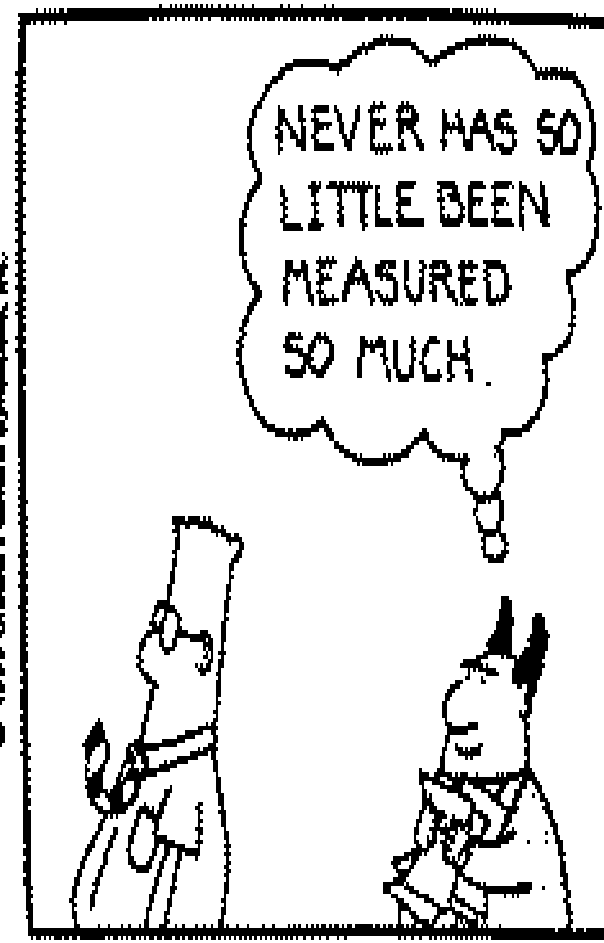
Core Issues	
Relevance	
Issue #1: Continued Need for program	<i>Assessment of the extent to which the program continues to address a demonstrable need and is responsive to the needs of Canadians</i>
Issue #2: Alignment with Government Priorities	<i>Assessment of the linkages between program objectives and (i) federal government priorities and (ii) departmental strategic outcomes</i>
Issue #3: Alignment with Federal Roles and Responsibilities	<i>Assessment of the role and responsibilities for the federal government in delivering the program</i>

- #1 Implied in the term *demonstrable need* is whether “private sector” opportunities exist (or have been displaced) or whether other orders of government may be better positioned.
- #2 Horizontal initiatives are a complication for an easy response to this otherwise straightforward issue.
- #3 A key issue in roles and responsibilities is federal jurisdiction and constitutional alignment.

Performance

Performance (effectiveness, efficiency and economy)	
Issue #4: Achievement of Expected Outcomes	<i>Assessment of progress toward expected outcomes (incl. immediate, intermediate and ultimate outcomes) with reference to performance targets and program reach, program design, including the linkage and contribution of outputs to outcomes</i>
Issue #5: Demonstration of Efficiency and Economy	<i>Assessment of resource utilization in relation to the production of outputs and progress toward expected outcomes</i>

- #4 – We see a blurring of the old formative/summative evaluation. Also apparent is the need to show a causal/attribution link (contribution) and a validation of program theory.
- #5 – This is the cost effectiveness issue linked specifically to **economy** (are we acquiring resources/inputs at the lowest cost?) and **efficiency** (are the outputs being produced at the lowest unit cost?). The **cost- effectiveness** question (cost per unit outcome) is implied in the term “progress toward expected outcomes.”



Translating TB issues into an effective evaluation matrix

- These five questions represent an abstract structure – evaluators need to translate these into concrete issues and questions pertinent to management needs.
- In general, the evaluation matrix must develop specific questions that support TB questions, **but...**
- The TB questions would rarely be posed on any evaluation in this format.
- Rephrase the TB questions to support reliable, valid, and concrete indicators that reflect the program context, goals, and implementation.



Focussing Evaluations

Because of resource constraints (time and money), and the increased demands imposed by effective triangulation, evaluators and managers will need to *focus* evaluations.

This focus will be driven by a risk assessment based on

- materiality and**
- strategic importance**

Risk assessment: strategic importance

Programs that are integral to the Department Agency priorities

Programs that, if they fail:

- Compromise the delivery of other programs
- Compromise the priorities of the department
- Cause social, economic, and political cost disproportionate to their magnitude

Programs that are constitutionally and legislatively required may be assigned a lower priority since the discretion on spending is limited

- Note that the Policy identifies these as requiring only an “administrative review”

The PAA level determines emphasis



Focus on relevance of activities and coherence of sub and sub-sub activities. Less emphasis on performance except as established by evaluations at lower levels

Key methods

- Literature reviews
- Expert interviews
- Senior management interviews
- Document reviews
- Performance (aggregation from evaluations of lower order PAA)

Emphasis on relevance and alignment to immediate level of PAA.

Key methods

- Limited literature review (operational, delivery, and outcomes of cognate programs)
- Management interviews
- Project level reporting (aggregation to sub and sub-sub activity level)
- Quantitative measures (surveys, administrative files, etc.)



Theory of change and Logic Models

Key ideas

- **Theory** explains the intervention and what outcomes are expected
- **Logic model** – two perspectives
 - explains the intervention (***causal logic***)
 - explains the organization of the intervention and how it integrates with broader objectives of government (***logistical logic***)
- **Performance measurement**



Causal logic models

- **Verbal** – explains the intervention and how it interacts with external events
- **Graphical** – presents a “picture” of the program
- **Abstract** (mathematical) – formalism that is most useful when quantitative data are available.

Causal logic models

Verbal models

National Child Benefit (NCB)

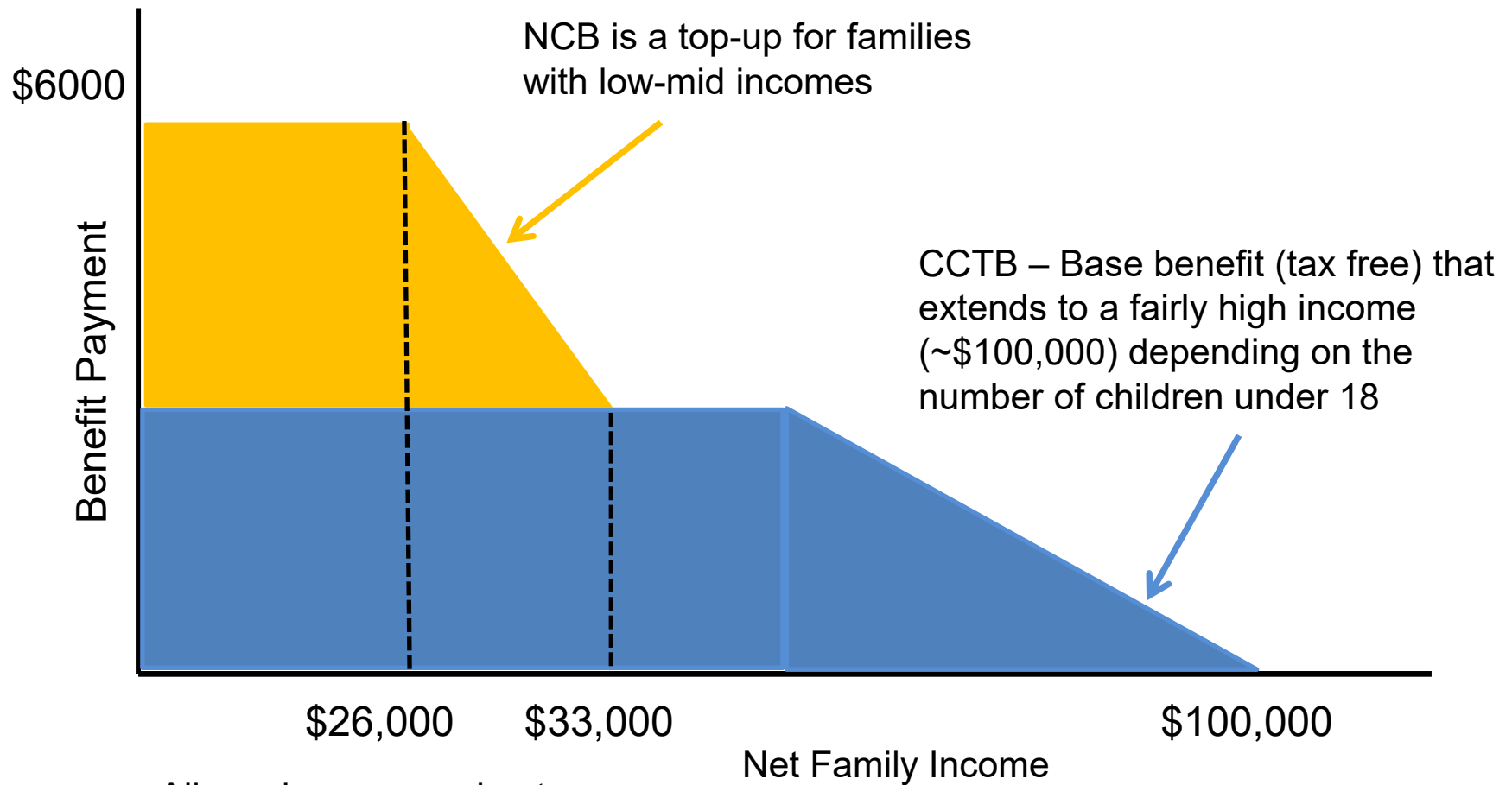
The NCB Initiative is a joint initiative of federal and provincial/territorial governments intended to help prevent and reduce the depth of child poverty, as well as promote attachment to the workforce by ensuring that families will always be better off as a result of working.

It does this through a cash benefit paid to low income families with children, a social assistance offset and various supplementary programs provided by provinces and territories.



National Child Benefit

(two children < 18)

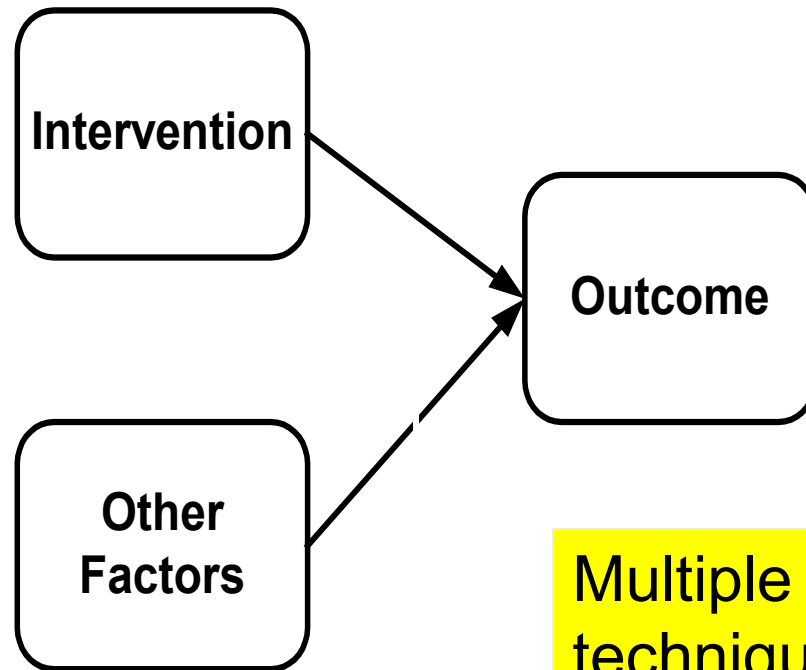


All numbers approximate



Causal logic models

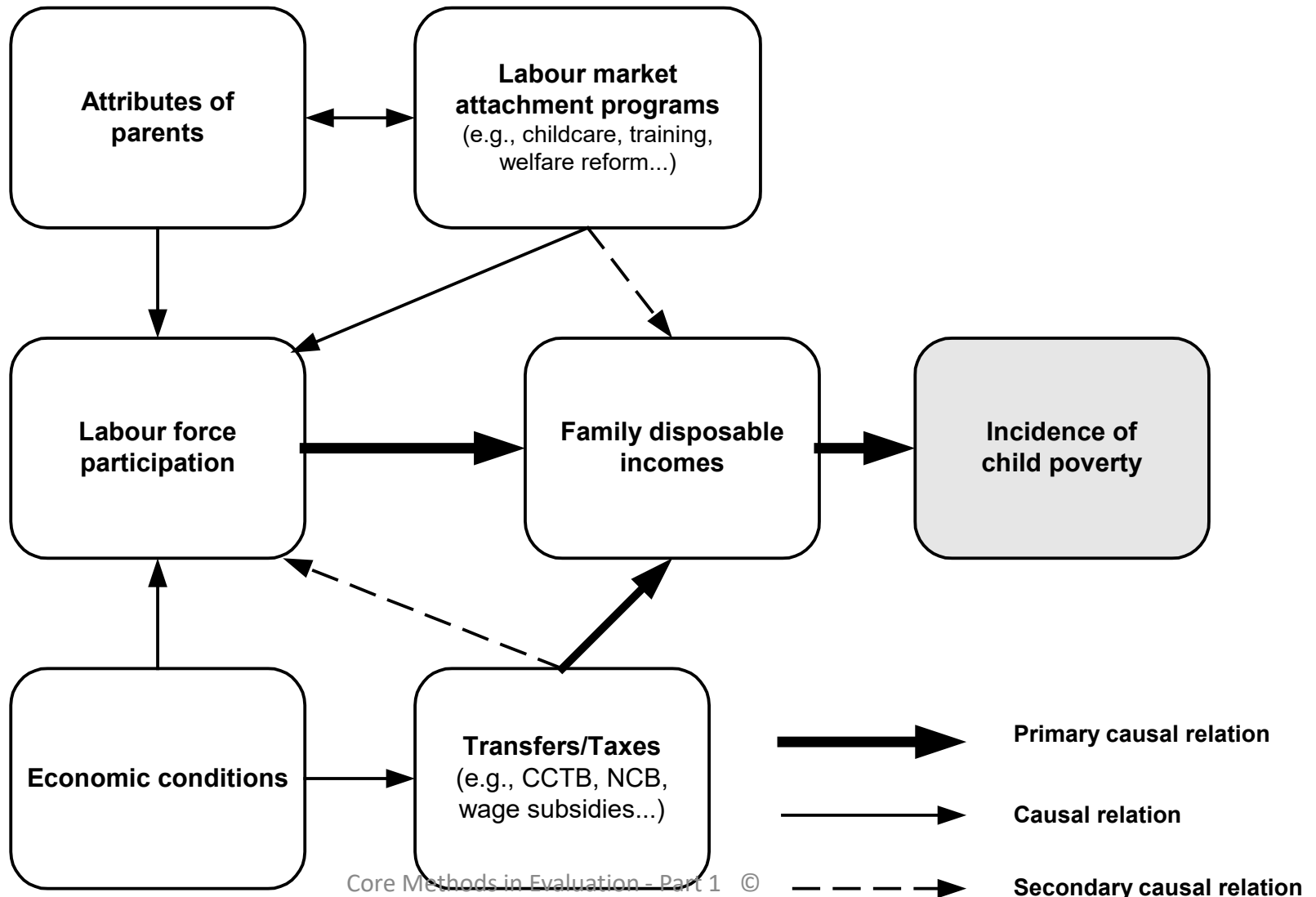
Graphical Models



The causal logic model clarifies the theory of how interventions produce outcomes.

Multiple methods and experimental techniques establish the relative importance of causes of changes in outcomes

Graphical logic for the National Child Benefit





Advantages and disadvantages of causal logic models

Advantages

- reveals inter-relationships among program elements.
- identifies confounding factors that reduce program outcomes.

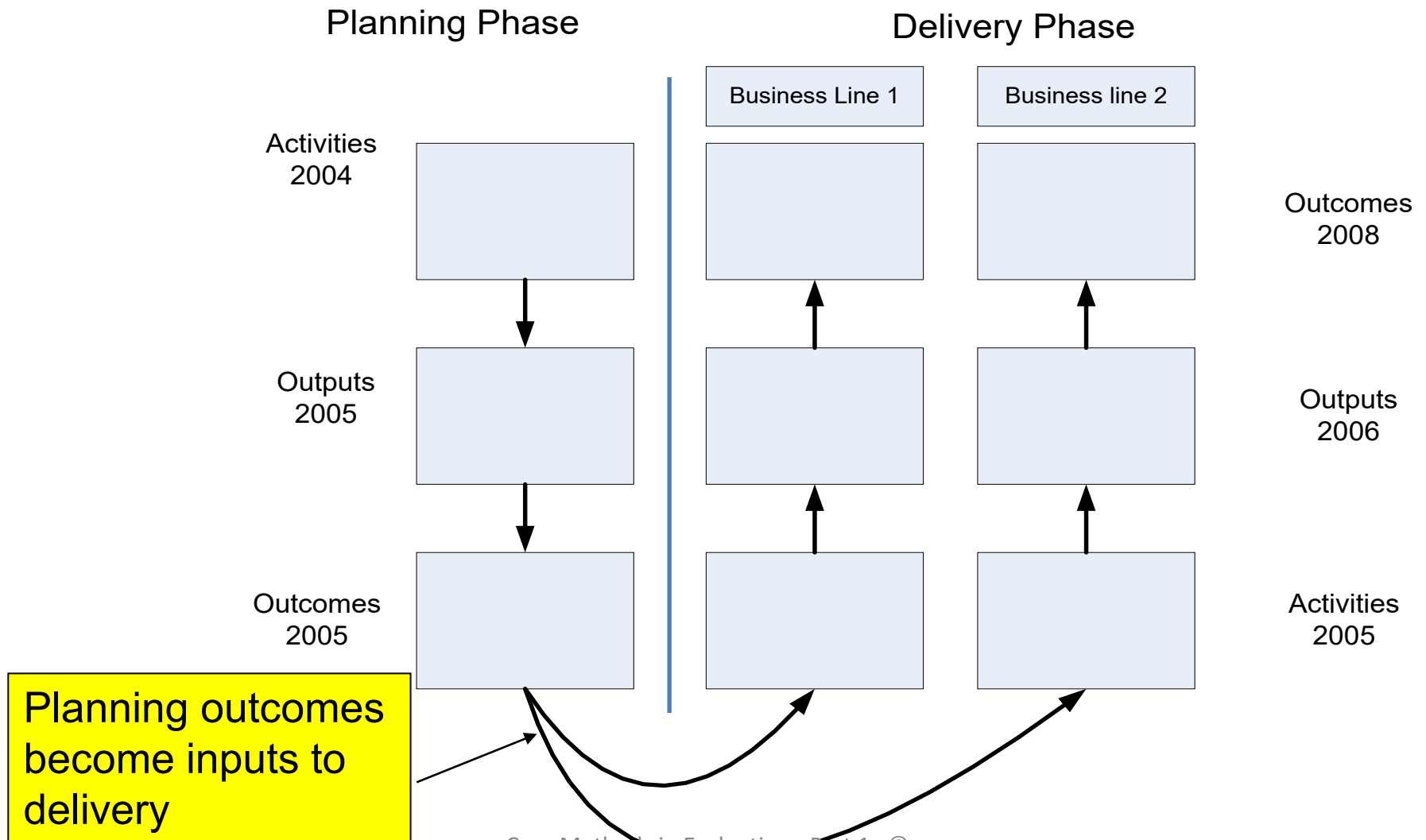
Disadvantages

- over-complication can impede understanding.
- abstract representations can confine communication.
- does not reveal resource use, reach or support other “oversight” requirements.

Implementation

- The conventional logic model offers no guidance on
 - Sequence of expected outputs (and therefore no insight in the timing of expected outcomes).
 - No clarity on immediate and intermediate outcomes.
 - Critical dependencies among program outputs and between program outputs and external processes.
- Implementation sequence is essential for developing sensible performance measurement.

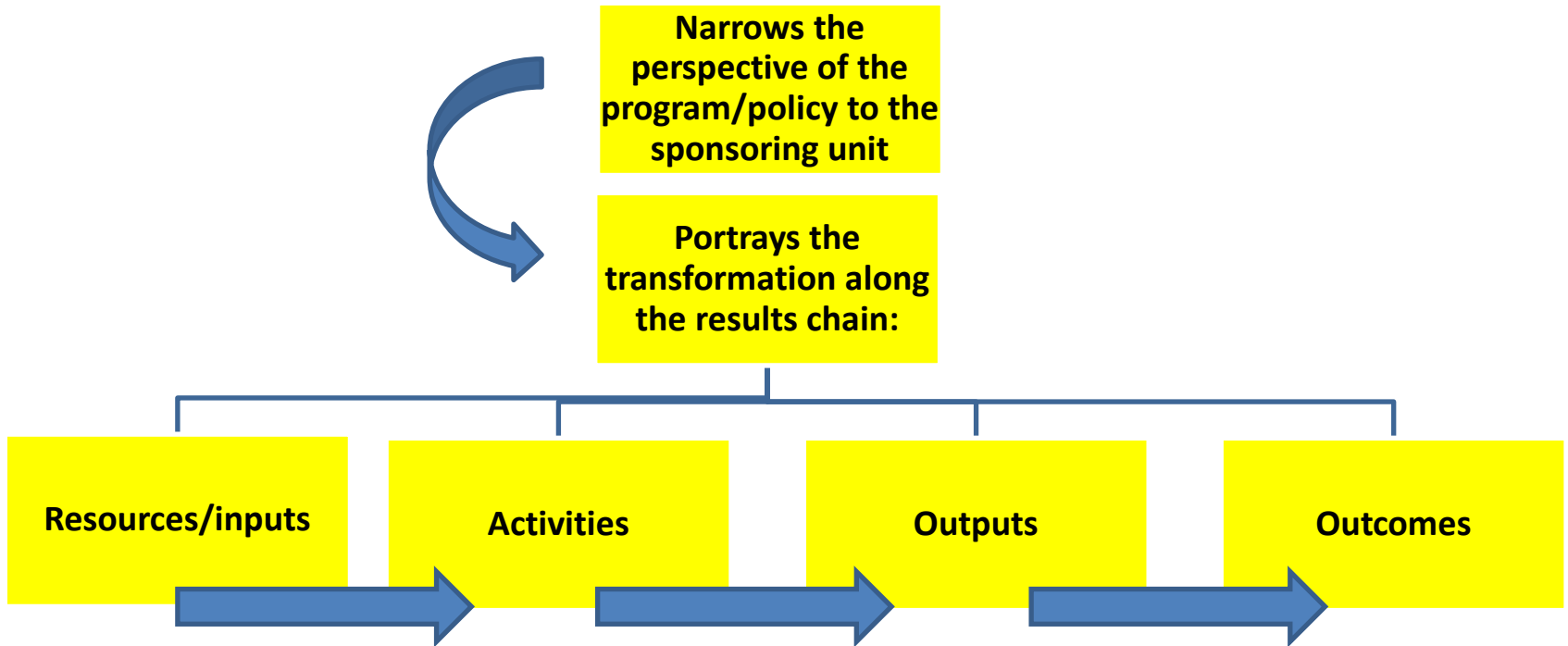
A logic model with time



Application to Horizontal Initiatives

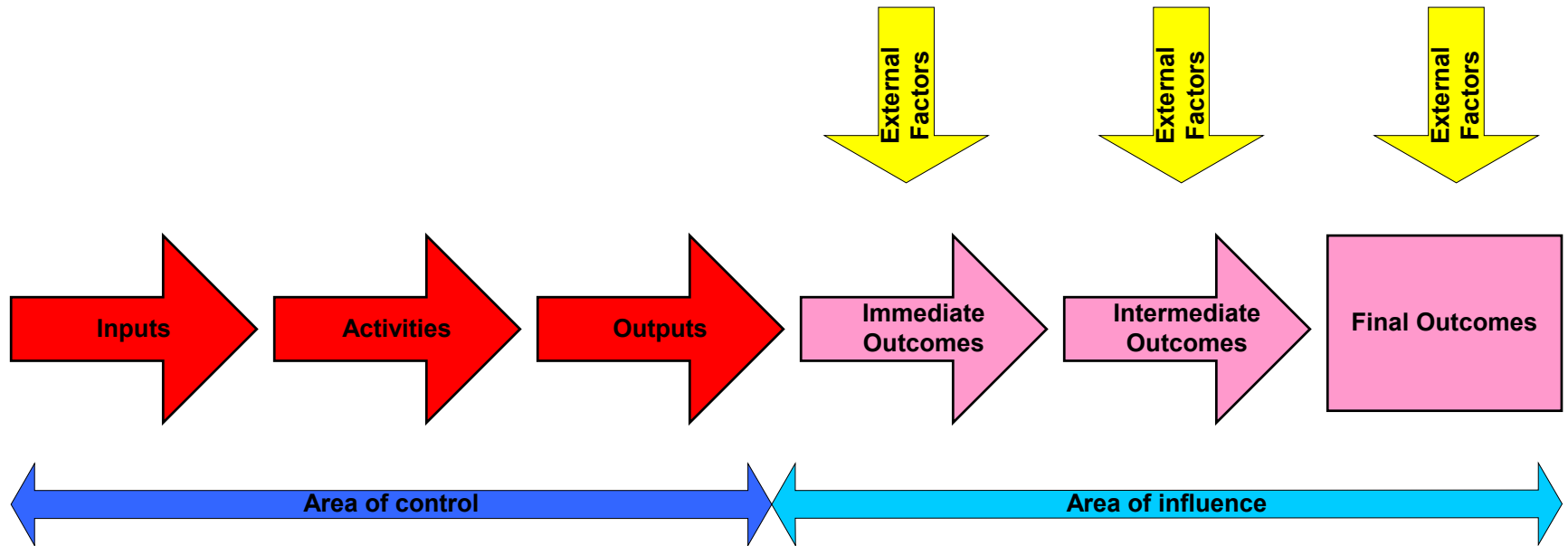
- The standard logic model confuses rather than clarifies.
- It fails to show how outputs from one partner process may be inputs to another partner process (critical path dependencies).
- The most important risk is failure to coordinate
- Critical path analysis and Gantt charts are useful “logic models” for complex initiatives
- The key logic requirement for the horizontal initiative is the “sequencing and mapping all activities/outputs/outcomes/costs.”

Logistics models



Logistical models

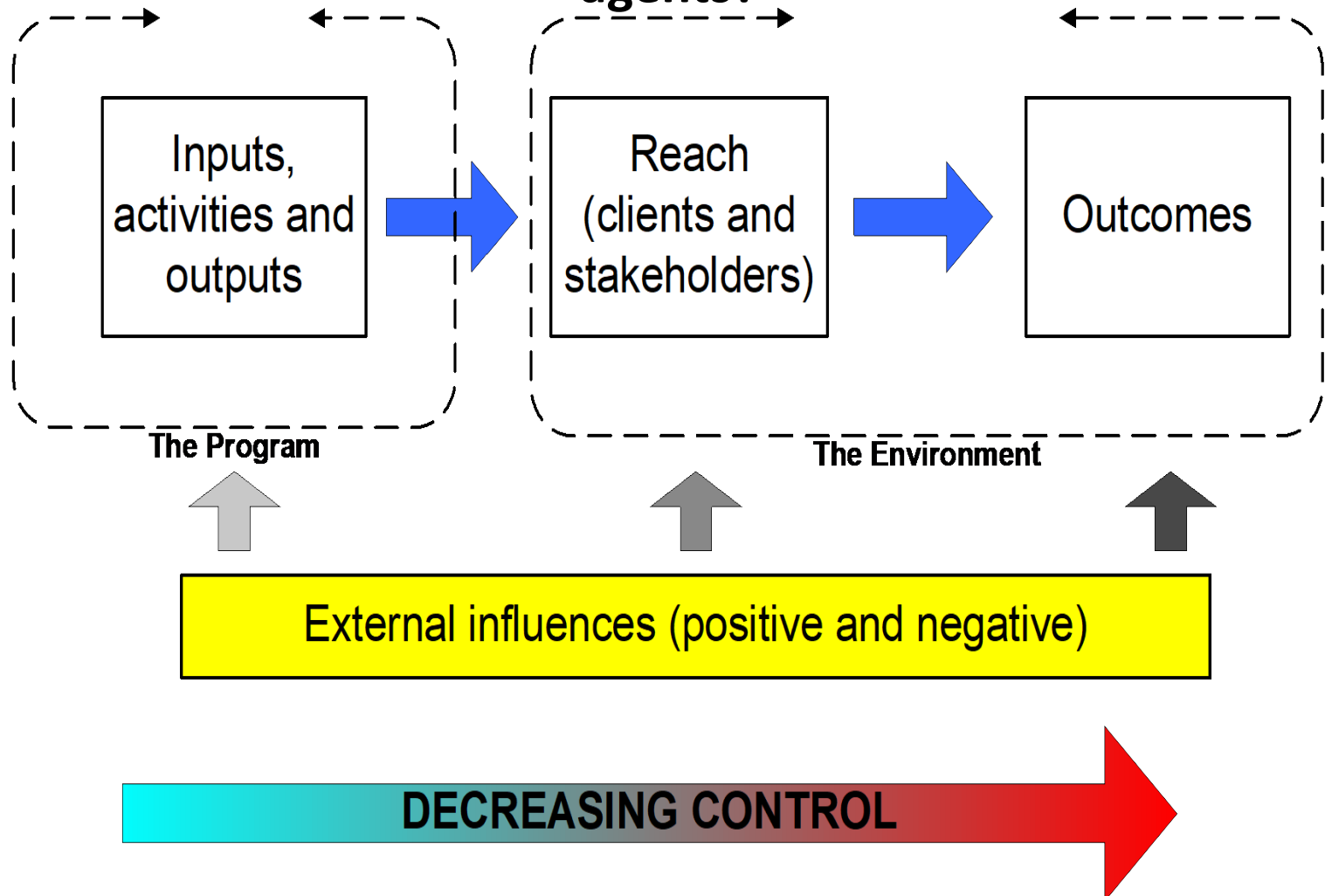
The results chain





Logistical models

Reach – who are the clients, stakeholders and delivery agents?



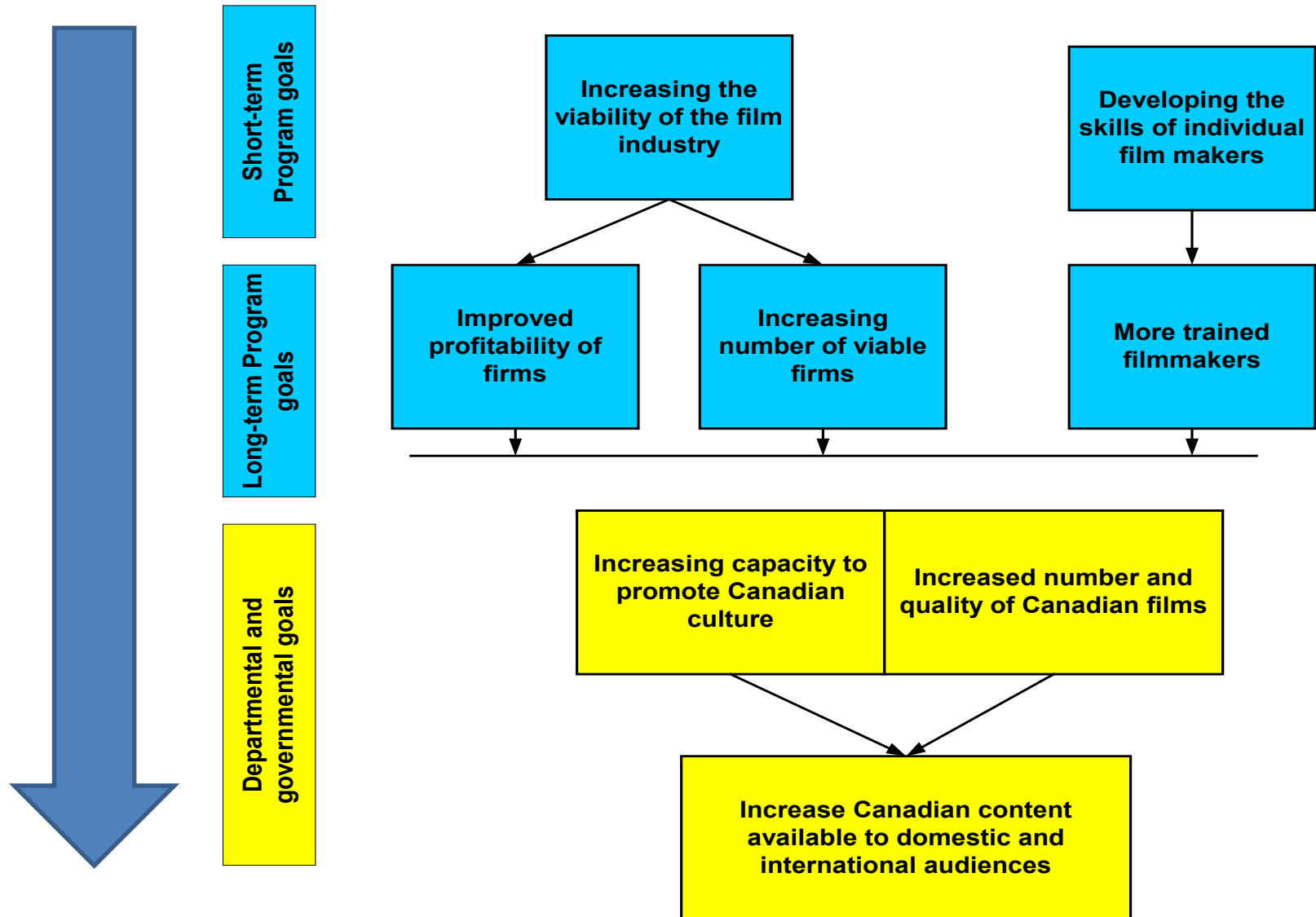


Logistical models



- An **output** is a product or service that is completely controlled by the program.
- An **outcome** is the change on the environment produced by outputs, mediated by other factors.
- **Long term outcomes** are the goals of the program
- **Immediate outcomes** (< 3 years) are the most important to track (why?)

Hypothetical Film Development Program

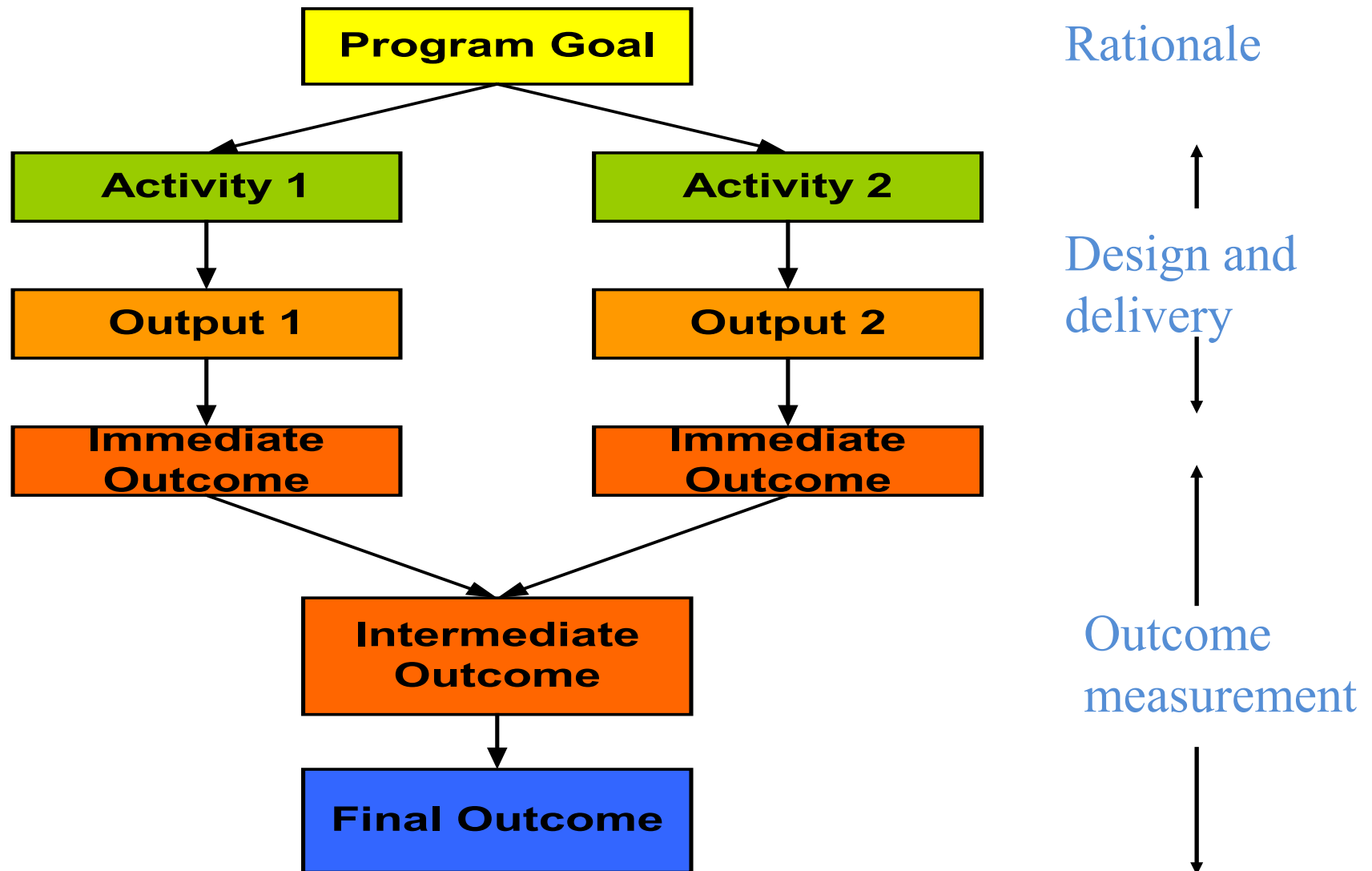


Performance must be measured along the results chain

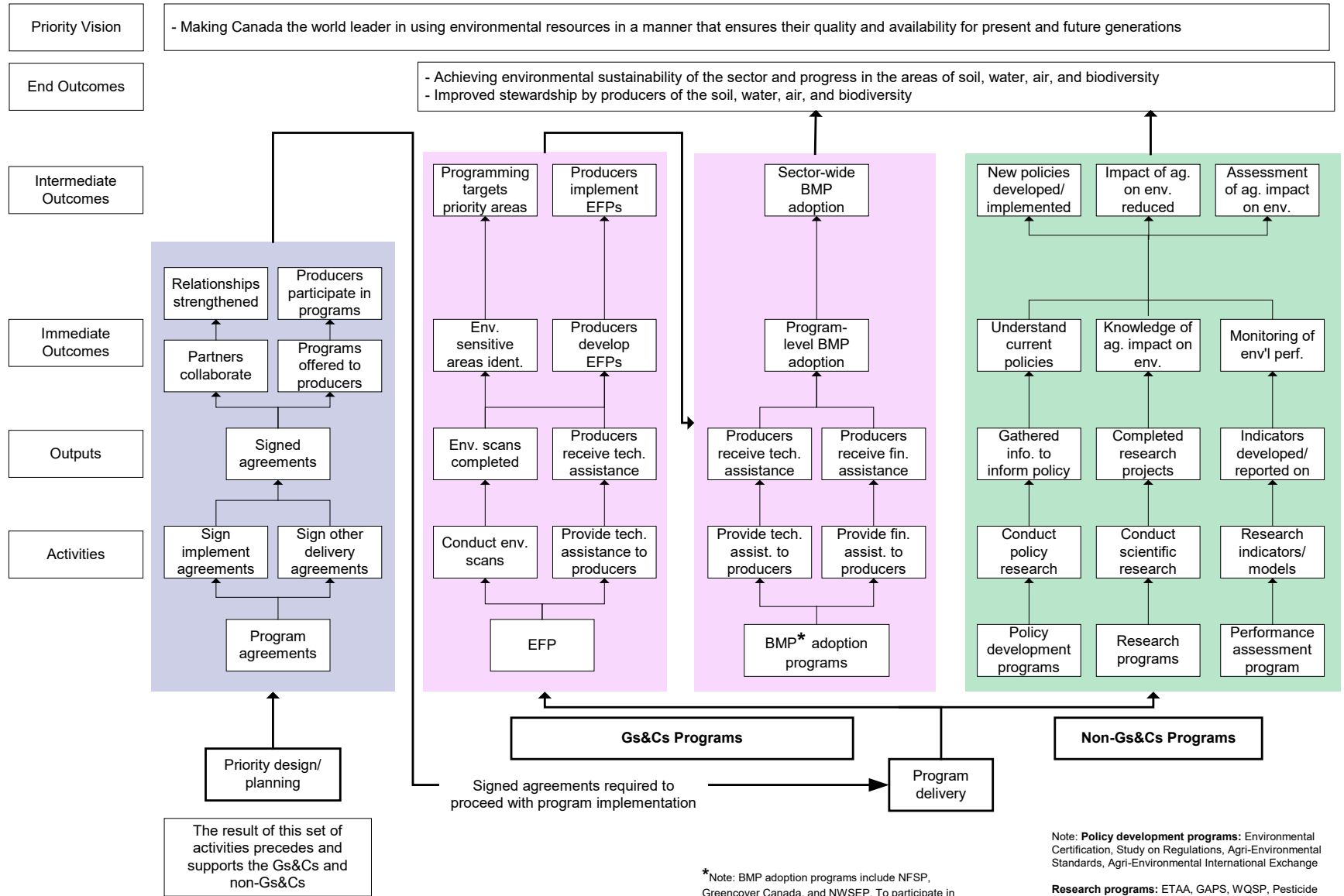
- ***rationale and need measures*** confirm that the intervention is needed/desired and that the the selected agency is most appropriate
- ***design measures*** examine the extent to which good theory supports the intervention.
- ***delivery measures*** assess organizational effectiveness of the intervention and whether the required outputs have been delivered, on time and where needed.
- ***immediate outcome measures*** determine immediate benefits for clients
- ***intermediate and long-term outcome measures*** track the effect of the outputs in producing sustained change to the environment, net of all external factors.



Performance measurement long the results chain



AAFC's APF Environmental Programs Organizational process and programming



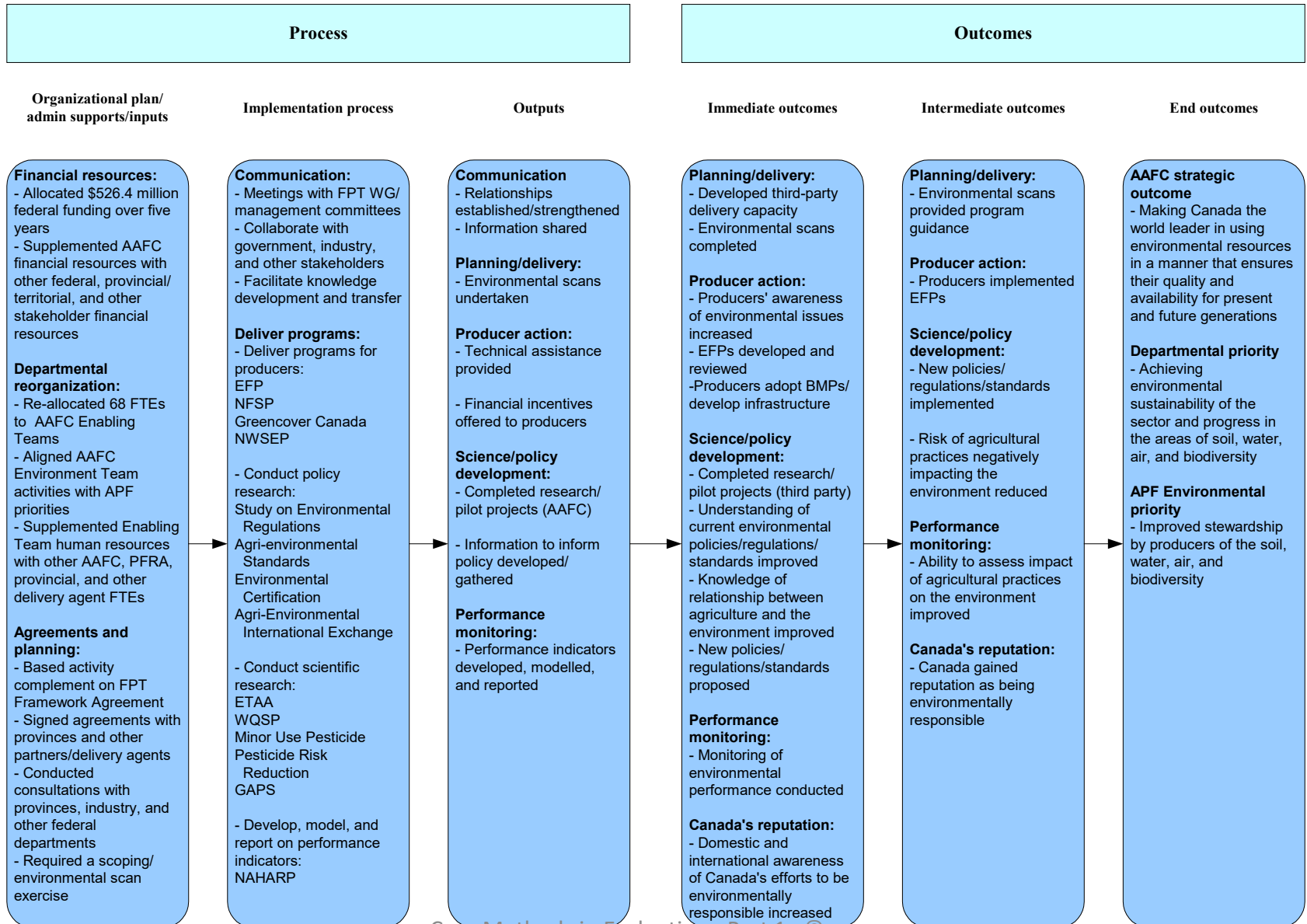
Note: **Policy development programs:** Environmental Certification, Study on Regulations, Agri-Environmental Standards, Agri-Environmental International Exchange

Research programs: ETAA, GAPS, WQSP, Pesticide Risk Reduction, Minor Use Pesticide

Performance assessment program: National Agri-Environmental Health Assessment and Reporting Program

*Note: BMP adoption programs include NFSP, Greencover Canada, and NWSEP. To participate in NFSP and some components of Greencover Canada, producers must have completed an EFP.

AAFC's APF Environmental Programs



From logic model to matrix

- Management shapes the purpose of evaluation by selecting the questions of interest
- The evaluation matrix guides the integration of data collection and analysis
- The matrix determines the success of the evaluation
- It shapes the direction and depth of analysis

IQIDM



Principles of evaluation matrix design (1)

- **Issues** need to align with the mission/goal of the program
 - High-level language is ok for the TB issues, but concrete and “grounded” plain language specifications are preferred
 - This translation will support measurement
 - Ensure that the service lines (program pillars) emerge at the issues level
- Do not replicate the TB structure ... it is too general
- Parsimony (a few focused issues) is preferred

Frameworks that reproduce the TB issues/questions reflect lazy work that sabotages mixed methods and effective triangulation (passing the buck)

Principles of evaluation matrix design (2)

- **Questions** are operational and specific to the program and service lines
 - Use the results chain and logic model to identify key delivery points/times/processes for *outputs*
 - Focus on immediate *outcomes*
- Questions align with indicators and data collection
- Rank questions within an importance level (H,M,L) to direct the allocation of evaluation resources.

Principles of evaluation matrix design (3)

- **Indicators** describe the information needed to answer the question
- Detailed descriptions support reliable and valid data collection
- Align indicators with source based on expected information content and quality

Example: What immediate outcomes (first five years) were expected at the program's inception?

- Client opinion (*poor*)
- Line management opinion (*slightly better*)
- Senior federal and provincial manager opinion (*even better*)

Principles of evaluation matrix design (4)

- **Data sources must align to each indicator**
- Detailed descriptions of sources must be specified to support efficient evaluations

Example: Senior federal and provincial managers' opinion

- Unspecified key informant interviews with a single guide (*poor*)
- Interview with federal agreement managers (n=3); Interview with ADM(s) (n=2); Interview with Provincial/territorial Agreement managers (n=13) (*better*)
- Align types of key informants to specific questions and create specialized interview guides for each class of key informant (*best*)

Principles of evaluation matrix design

- **Methodology** explanation adds important detail for each data source and indicator
- The plans for integrating data and lines of evidence starts with the matrix

**KISS – Keep it simple
and sophisticated**

Many evaluation matrices are bloated and repetitive with redundant questions that fail to direct the collection of strategic data

Case example 4 – Horizontal Summative Evaluation of the Government of Canada's Investment in the 2010 Olympic and Paralympic Winter Games

Immediate outcomes

- Leverage 2010 Games to advance existing federal priorities
- Positive exposure and heightened recognition of the Government of Canada as a key partner in the 2010 Games
- Successful delivery of the mandated essential federal services.

Intermediate outcomes

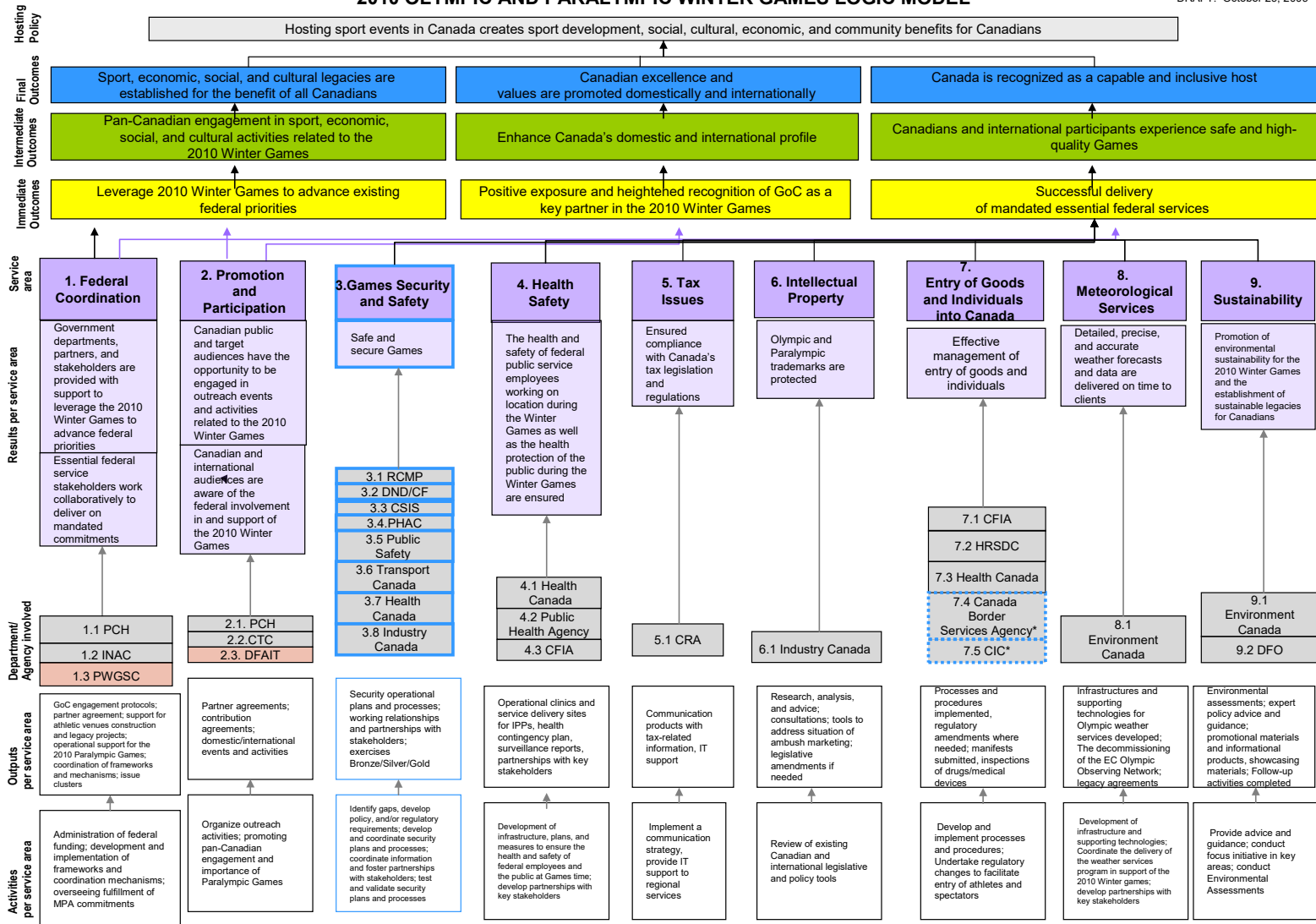
- Pan-Canadian engagement in sport, economic, social, and cultural activities related to the 2010 Winter Games
- Enhance Canada's domestic and international profile
- Canadians and international participants experience safe and high-quality Games

Final outcomes

- Sport, economic, social, and cultural legacies are established for the benefit of Canadians
- Canadian excellence and values are promoted domestically and internationally
- Canada is recognized as a capable and inclusive host.

2010 OLYMPIC AND PARALYMPIC WINTER GAMES LOGIC MODEL

DRAFT: October 29, 2009



* This Department/Agency, although they contribute directly to the result of service area "Entry of goods and individuals into Canada", will report and be evaluated under service area #3 Games Security and Safety as they have significant contribution to the result of this area.

1



Main methods used in 2010 Olympics

- Focus on the role of the Federal Secretariat and the contributions of:
 - 15 federal departments provide services deemed essential for conducting successful Games
 - Auxiliary services
 - PCH received support for opening ceremonies
 - INAC received funding to ensure legacies, benefits and participation of first nations
- Security (separate evaluation conducted by RCMP)
- Eight other studies comprised the evaluation
- Polling used to track “national pride”
- Media analysis complemented quantitative polling
- Interviews (directed to specific classes of respondents).

Insert

[..\..\..\GOODWILL\CES\CES2013\NCC Jan 24\Table A2.docx](#)

Day 1

Foundations

**Understanding and assessing
relevance**

Afternoon
January 31, 2013



Foundations of Policy Development

- Public policy rests responding to deviations from some norm.
- Norms derive from social, religious, cultural, ethical or economic principles.
- The decision to change a “natural outcome” rests on a comparison of the intervention costs compared to the intervention benefits.
- Economists use the norm of a “competitive economy” – the prices (including wages, interest, profit, etc.) that would ensure in a hypothetical world of perfect competition.
- The economic model leads to a robust structure for modelling policy.
- In the last year the theory of macro-economic policy has come under substantial criticism.
- Micro-economics, however, remains largely intact.

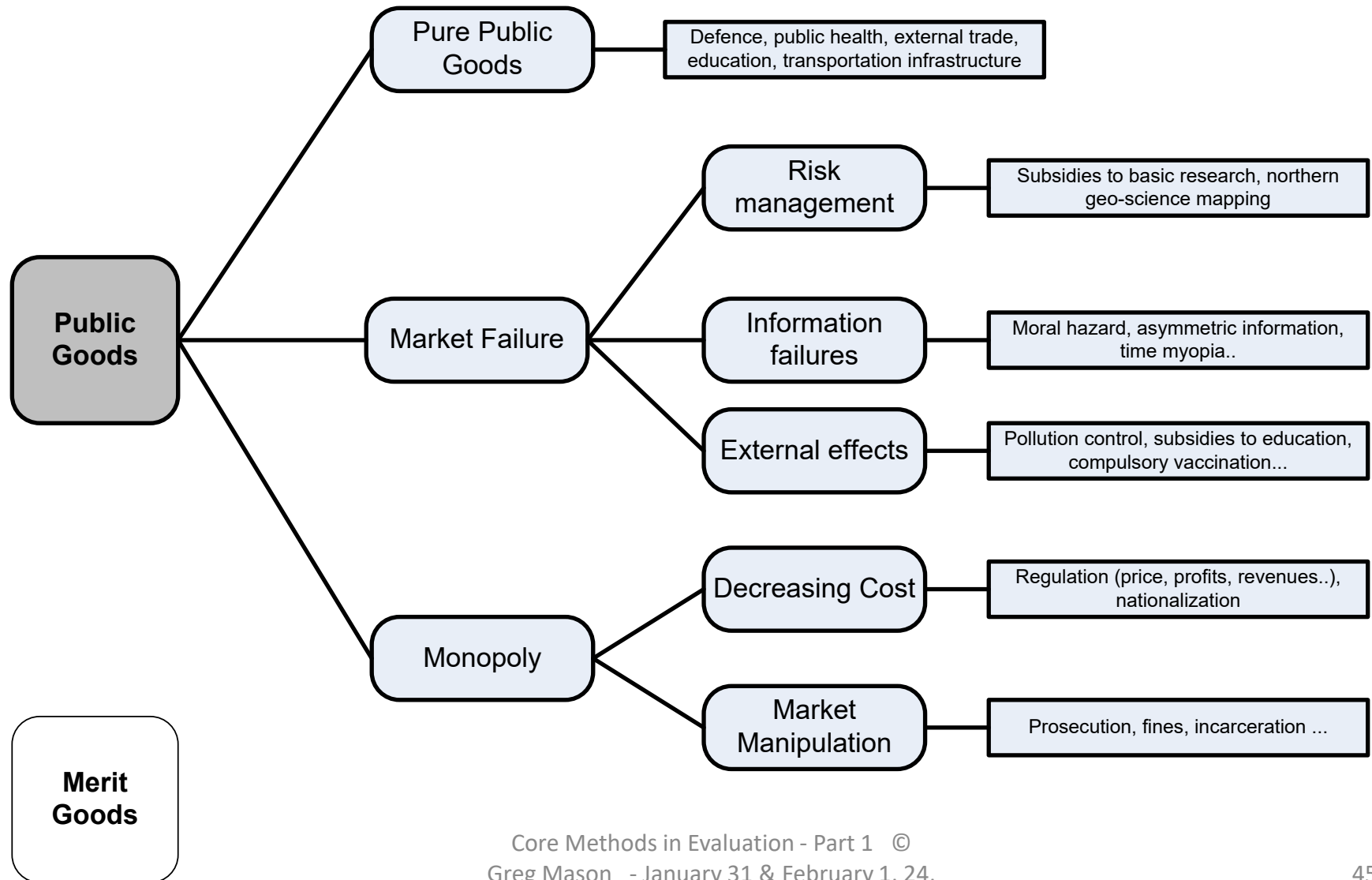
Why does government exist?

Three main rationales for public sector action:

- Market failure (consumer ignorance of mortgages, pollution)
- Externalities (public goods and bads)
- Distributional unfairness (poverty)

- 1. Market failure** typically evokes a regulatory response (e.g., consumer education, fair lending laws, securities regulation).
- 2. Public goods** encourage government to supplement private sector provision of a good or services (e.g., subsidization of crop insurance, subsidization of vaccines, public education).
- 3. Distributional fairness** can result in regulatory, direct provision of a service, or direct cash transfer
 - Laws regarding usury, anti-discrimination legislation
 - Public housing
 - National child benefit, progressive tax, GST rebate for lower income households

Government provided goods and services





Definition of government initiatives

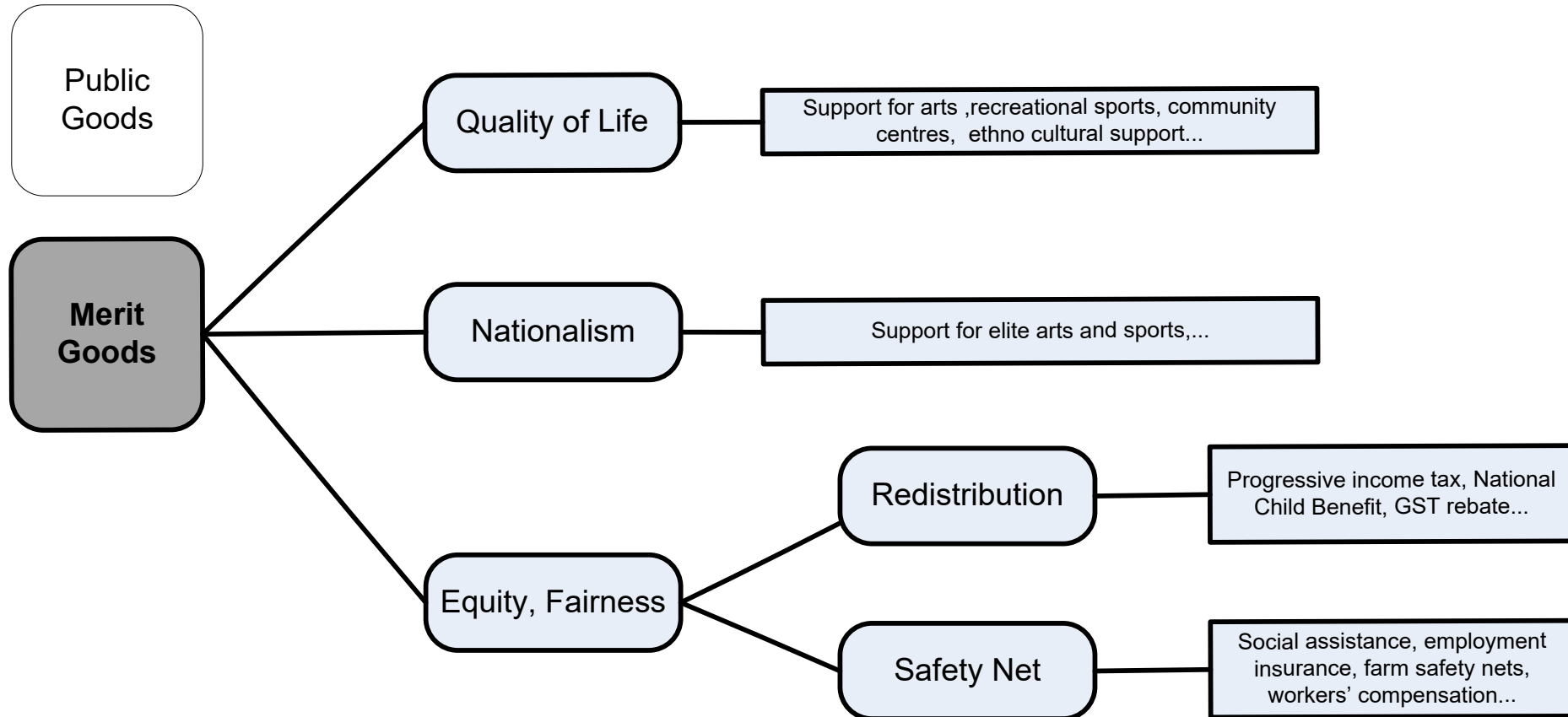
- **Social marketing to promote a goal** (articulation of goal or intent; guidance on preferred behaviour)
- **Expenditures on goods and services**
 - *Direct resource commitments* on goods (public housing, vaccination)
 - *Direct resource commitments* on services (consumer information, training)
 - *Tax expenditures* (tax deductions and credits awarded to citizens and businesses to behave, spend, invest, etc.)
 - *Grants/contributions/contracts* to third parties to perform services
- **Legislation** is a general framework for how citizens conduct themselves (smoking bans, criminal code) and requires political assent.
- **Regulation** modifies elements of legislation (changes to the speed limit) and can be completed by administrative fiat.



Information Failure

- **Moral hazard**
 - Market participants alter their behaviour in response to the divergence of public and private costs
 - Taxes/subsidies cause market participants to purchase/sell less/more than would have occurred with prices equal to the marginal cost
- **Asymmetry of information**
 - Sellers are typically more informed than buyers
 - Prisoners paradox - information lack produces sub-optimal outcomes
- **Uncertainty** about other players reactions causes poor decisions
 - Nash equilibrium exists when I account for your probable reaction to my choices. Equilibrium exists when we have all adjusted and readjusted to each others choices/decisions.

Government provided goods and services



Distributive justice as a merit good

- Efficiency is not the only goal for government

One can judge a society by how it takes care of its weakest.

Daniel Moynihan (US Senate)

- **Pareto rule** – make only those changes that benefit at least one person and make no one else worse off.
 - Politically impractical
 - Tolerates a society where one person has everything
- **Kaldor-Hicks** rule – make a change if, in principle, those who benefit could compensate those that lose
 - Typically government uses tax rules and subsidies to effect compensation.

Recent research shows that ideas of equality and fair distribution become settled for most people by the age of 10. Strong evidence exists that humans develop altruistic instincts early.

“Share and share alike”, *Nature*. 454(28) Aug 2008

Core Methods in Evaluation - Part 1 ©

Greg Mason - January 31 & February 1, 24,

2013

Fiscal Federalism: the Canadian overlay

- **Fiscal federalism**
 - Allocation taxation powers among orders of government (federal – provincial – territorial)
 - Flows of financial support equalization of access to core services (e.g., Canada Health Act- 1985)
 - Evolution of service delivery
 - Increased devolution to lower orders of government
 - Outsourcing (G&Cs, privatization)
 - Insourcing (regulation, nationalization)
- **Key legislation and antecedents**
 - The Constitution Acts (1967 and 1982)
 - Rowell-Sirois Commission (1940) – (sharing federal-provincial responsibilities)

Typical documents – information potential

Document type	Information content – potential role in the evaluation [1 = qualitative 2 = quantitative]
Foundation documents (TB Subs, MCs, policy background)	<ul style="list-style-type: none"> • Program rationale and relevance [1] • Program origins [1] • Authority (financial, governance) [1] • Outcomes [1,2] • Targets [1]
Performance reports (G&Cs)	<ul style="list-style-type: none"> • Outputs and outcomes [2] • “Thick” descriptions (implementation, outputs, outcomes) [1]
Audits and evaluations	<ul style="list-style-type: none"> • Program history [1] • Benchmark for costs, implementation outputs, outcomes [2]
Applications (G&Cs)	<ul style="list-style-type: none"> • Applicant/client attributes [1,2]



Typical administrative files – information potential

Document type	Information content – potential role in the evaluation [<i>1 = qualitative; 2 = quantitative</i>]
Management files (meeting minutes, HR records, etc.)	<ul style="list-style-type: none">• Number [2] and type of employee [1, 2]• Minutes of meetings to<ul style="list-style-type: none">– describe implementation [1 and 2]– participation of partners: number [2] and type [1]– implementation timing [2] and processes [1]
Financial records	<ul style="list-style-type: none">• Payments (individual and aggregate) [2]• Distribution and fairness [2]• Payment timing and delay [2]
Client services	<ul style="list-style-type: none">• Services delivered [1, 2]• Participation in program [2]• Sample frame to support survey [2]

Typical literature reviews – information potential

Review type	Information content – potential role in the evaluation [1 = qualitative 2 = quantitative]
Scan	<ul style="list-style-type: none"> • Top line summary [1] • Program context [1] • Implementation context [1]
Integrated review	<ul style="list-style-type: none"> • Program “arc” (history, context, and evolution) [1] • Theory of change [1]
Integrated literature and expert review	<ul style="list-style-type: none"> • Program “arc” (history and evolution) [1] • Theory of change [1] • Theory of context and program evolution [1]

Case example

- Literature on Labour Market Information for Aboriginal Youth
- [F:\PROJECTS\CANADA\HRDC\038\(AboriginalLMI\)\REPORTS\docFinal Aboriginal LMI V4.doc](F:\PROJECTS\CANADA\HRDC\038(AboriginalLMI)\REPORTS\docFinal Aboriginal LMI V4.doc)

Day 2

Quantitative and qualitative data collection and analysis

Quantitative Methods

Morning
February 1, 2013



Data collection methods

Quantitative

- Questionnaires (large sample, self-administered)
- Structured interview (large sample, fixed response with interviewer)
- Observations (Structured)
- Content analysis of documents
- Administrative files (program activity, client activity, output/outcome counts, financial data)

Qualitative

- Participant observation (observer takes unstructured notes)
- Observation (semi-structured and unstructured)
- Focused interview (key informant)
- In-depth interview
- Oral history
- Content analysis of documents

In general, qualitative research is more labour intensive (costly) than quantitative research.



Quantitative Research

- **Unit of analysis** aligned to the 1) program target focus
 - Individuals
 - Families/households
 - Firms
 - Organizations
 - ...
- **Unit of analysis** aligned to 2) the program delivery focus
 - Managers
 - Organizations
 - ...

Key idea: Quantitative methods rely on “counting” similar units

"use of standardised measures so that the varying perspectives and experiences of people can be fit into a limited number of predetermined response categories to which numbers are assigned" (Patton, 2001, p.14).



**87% OF THE 56% WHO COMPLETED MORE
THAN 23% OF THE SURVEY THOUGHT IT
WAS A WASTE OF TIME**

Typical large probability sample surveys reviews – information potential

Survey type	Information content – potential role in the evaluation [1 = qualitative 2 = quantitative]
Interviewer mediated	<ul style="list-style-type: none"> • Respondent self-report <ul style="list-style-type: none"> – Fixed response – number/category [1 and 2] – Verbatim [1] • Interviewer probes [1 and 2] <ul style="list-style-type: none"> • Interviewer-respondent interaction creates a complex qualitative data field [1] • Potential to increase reliability and validity [1 leads to insight on quantitative results 2, if data are presented to interviewees • And decrease reliability and validity [interviewer knowledge and skill paramount]
Self-completed	<ul style="list-style-type: none"> • Respondent self-report <ul style="list-style-type: none"> – Fixed response – number/category [1 and 2] – Verbatim [1 → 2 on coding]

Sample surveys represent a blend of qualitative and quantitative methods

1. Design phase

- Literature
- Standard scales
- Prior surveys
- Expert interviews
- Focus groups

2. Pretest

- Expert assessment (detect hesitancy)
- Active probes to secure meaning
- Follow-up debrief

3. Data collection

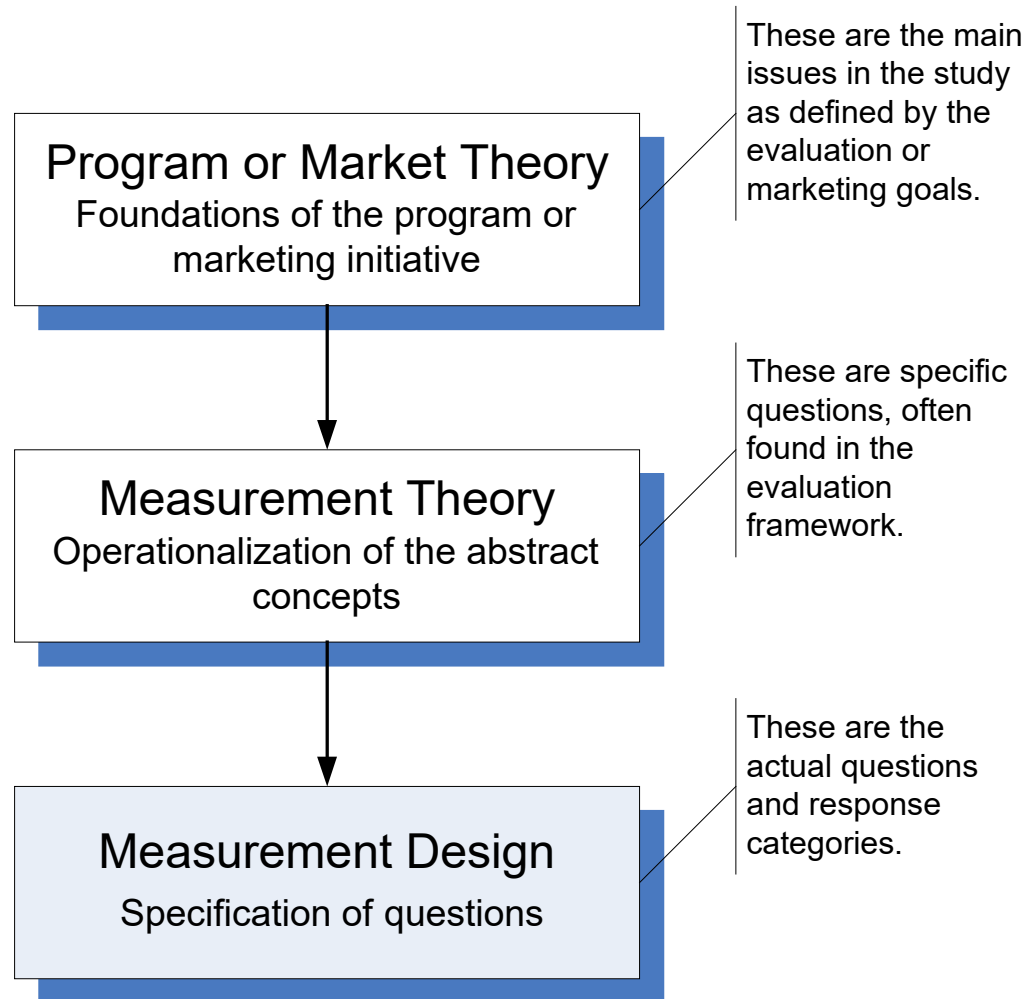
- Verbatim entry

4. Analysis

- Coding and categorization

5. Reaction to results by experts and KIs

Questions 101 – The basic structure



Question phrasing ‘rules’

- **Set wording to the respondent** — Make sure that jargon and acronyms will be understood by the respondent.
 - Use a Flesch-Kincaid Grade 8 level for the general population and Grade 11 for civil servants and professionals.
 - Use technical jargon with specialized audiences to communicate that you understand issues.
- **Short sentences** in the active voice work best.
- **Balance alternatives**
 - “Some people support sending Canadian troops to Afghanistan; others do not. What do you think?”
 - Instead of “What do you think of sending Canadian troops to Afghanistan?”

Question phrasing ‘rules’ (cont’d)

- Avoiding leading the respondent
 - *“In order to balance the budget, should government reduce spending on emerging artists or raise taxes on the rich?”* is clearly biased.
- Instead, use two questions:
 - *Do you agree or disagree that the budget should be balanced?*
 - *[If agree] Should government reduce spending on artists or raise taxes on high income residents?*
- Or better still, use the same phrasing as the previous slide for the second question:
 - *To balance the budget, some advocate increasing taxes on high income residents, while others advocate reducing spending on artists. What do you think? (Rotate)*

For stylistic reasons, one might be tempted to replace the second use of “advocate” with a word like “support,” but this makes the question unbalanced since “advocacy” is a stronger word than “support” unless we rotate the question.

Question phrasing 'rules' (cont'd)

- **Ground and focus the question** (See section on cognitive interviewing)
- **Avoid hypothetical questions** that are improperly framed. Special techniques (discrete choice and conjoint) create a structure for these questions (see below).
- **Don't know, no opinion, neutral, and not applicable are different**
 - **Don't know** – respondent is unfamiliar with the topic and cannot form an opinion
 - **No opinion** – respondent knows the topic but is disinterested in any alternative
 - **Neutral** – respondent knows the issue, but has adopted a position within the extremes
 - **Not applicable** – respondent is not eligible to respond, regardless of their knowledge or strength of position.

Framing questions and cognitive interviewing

Classic study (Belson, 1981)

In a face-to-face interview, respondents were asked to agree to disagree with a series of statements such as “television shows are too violent for children.” After carefully recording the responses, respondents were approached the day after and “debriefed” about the survey. Interviewers asked the respondents what they meant by “television shows,” “too violent,” and “children”. Belson discovered that these terms meant different things to different people, and he detected distinct meanings.

- Television show meant prime time to some, and all times to others.
- Children meant under 6, under 12, and under 18 depending on the respondent.
- Too violent had meaning specific for each individual.

With three meanings of child and two meanings of TV show, there are six questions being asked, let alone the infinite shades of “too violent.”

If all the problems of question wording could be traced to a single source, their common origin would probably prove to be in taking too much for granted.

S. Payne, *The Art of Asking Questions*, 1951

Cognitive interviews

- **Key challenge** - to phrase the question so that everyone understand what the research means.
- **Requires extensive pretesting** using cognitive interviewing
- **Cognitive interviewing uses “*think-alouds*” and “*probes*.”**
 - *Think-alouds* allow the respondent to collect their thoughts verbally

“In the last six months have you been to the dentist” <yes> “Please describe the times you say the dentist, <*let me see, I had my teeth cleaned six months ago and last week has to go for a repair on a chipped tooth.. Oh yes, I had a tooth ache last April*>
 - *Probes* direct specific questions

“Tell me about the last time you went to the dentist.” “Was this for a checkup or to deal with a problem?” “Why are you certain that this last visit was in June?”

Grounding the question

- **Cognitive interviewing** assesses response variation in a specific question.
 - **a pretest technique** for mapping response variation and error (deviation from the intended meaning).
- **Two models:**
 - **Ericsson-Simon model** (1980)
 - presumes people can remember why they responded to a question the way they did.
 - experimental evidence shows that this works when the recall task involves verbal information (as opposed to non-verbal/spatial information), is novel, endures for a period of time, and has happened recently.
 - It also works when subjects are asked to describe “what” as opposed to “why” they did something.
 - **Task analysis** (Tourangeau, 1984).
 - questions are processed according to a protocol and answers are provided within a specific values context.

Task analysis contexts

Question-answer processing

- Cannell et al. (1981)
 - comprehension
 - decision/retrieval/organization of data
 - response evaluation (filtering)
 - response output
- Martin (1983)
 - giving meaning to the question
 - searching for relevant data
 - formulating a judgment
- Tourangeau (1984)
 - comprehension
 - retrieval
 - judgement
 - response

Framing

- Many questionnaires make excessive demands on memory.
- The term “recall bias” is misleading or oversimplified when it actually means “collecting really bad data.”
- Framing practices
 - Use introductions and questions to set the stage.
 - Send a letter in advance explaining the survey and reminding the respondent of key dates (e.g., *Our records show you were a patient at Acme Cardiac and Rotor Rooter Unit four months before you died.*)
 - Avoid asking detailed questions about events or states in the past.

1. *Four years ago, in 2003, how much money did you normally make at your job in a week, before taxes? If you did not have a job, please write “zero” or the number “0”.*

Amount made per week _____

Advanced issues: Attributes of a good response scale

- Respondents and researchers need to accept that respondents' feelings, perceptions, and judgements can be described by numerical, semantic, and physical analogues.
- Respondents must share the same interpretation of the scale.
- A scale must discriminate among levels.
- A scale value must mean something. "Very good" must mean something different (better) than "average".

Example: (telephone or mail)

On a scale of 0 to 10, where 0 is poor and 10 is excellent, what is your overall opinion of the program?

Example: (mail)

Using the scale below, please rate your last experience with the emergency room.

-2 -1 0 1 2

(Very Negative) (Very Positive)

Magnitude scales

Advantages

- Identify a middle position (0 – 10)
- Linear with even steps
- Supports statistical measures of central tendency and variance

Disadvantages

- Need to translate subjective concepts into a number
- Linear scales may not capture intensity at the extremes
- Easy to overuse
- Automatic massing at the midpoint

Example: (telephone or mail)

On a scale of 0 to 10, where 0 is poor and 10 is excellent, what is your overall opinion of the program?

Example: (mail)

Using the scale below, please rate your last experience with the emergency room.



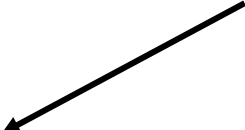
Two common biases

- *Inter-item contamination*

Qa. In your view, is AIDS a threat to someone who is heterosexual and not a drug user?

Qb. Is the government providing sufficient funding to basic research in health?

Qa Contaminates Qb



- *Social desirability bias*

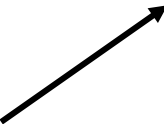
Qc. Have you heard of the XYZ program?

Qd. In order to assess how well we are promoting our services, please tell me whether you have heard of the XYZ program.

Challenges



Shifts blame and allows
someone to admit ignorance



Negative framing

Scenario 1

Imagine that Canada is preparing for the outbreak of an unusual disease that is expected to kill 600 people. Two alternative programs have been proposed on a survey of 100 people with the following outcomes: *[respondent numbers in brackets]*

- Program A 200 people will certainly be saved. *[72 favoured this one]*
- Program B There is a 33% chance that 600 will be saved and a 67% chance that no one will be saved. *[28 favoured this one]*

Scenario 2

Same basic scenario as above, but with the following two program alternatives:

- Program C 400 people will certainly die. *[22 favoured this]*
- Program D There is a 33% chance of 0 deaths and a 67% chance of 600 deaths. *[78 favoured this]*

Negative
framing
affects
response

Framing is neither good nor bad, but a feature of a linear structure sequence to a conversation.



Yes Prime Minister

Sir Humphrey teaches questionnaire design

Classic British TV comedy *Yes Prime Minister* has important lessons for those who want to interpret questionnaire data.

This clip shows two civil servants discussing a policy suggestion. Bernard Woolley, who we see first, thinks the public are in favour of the policy – the minister has had an opinion poll done. Senior civil servant, Sir Humphrey Appleby sets him straight.

Fans of cognitive biases, note that Sir Humphrey uses at least three in his illustration of a biased questionnaire: framing, priming, and acquiescence bias.

This example is exaggerated, but the moral still holds: questionnaires can be designed to encourage the answers you want. People's opinions are not objective facts like their height and weight. They change, depending on the context and on how they are asked.



[Yes-Minister-SUrvey.wmv](#)

Insert -

..\..\..\HERITAGE\043(2010 Olympics)\REPORTS\04
Phase 3 Reports\03 media analysis and POR
report\rptMedia analysis and POR v7.pdf

..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\history
e.pdf

..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\samples
ize e.pdf

..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\telepho
ne e.pdf

Day 2

Quantitative and qualitative data collection and analysis

Qualitative Methods

Afternoon
February 1, 2013



Qualitative Research

- Reliability and validity in quantitative research depends on instrument construction that is aligned to the hypotheses.
- Comprises any data that cannot be counted and processed statistically
- Common manifestation in evaluation are
 - Interviews
 - Focus groups
 - Case studies
- Key challenge is selecting subjects (as opposed to sampling) for their information value.

“the researcher is the instrument” (Patton, 2001, p. 14).

Reliability

- a. Many qualitative researchers argue that without validity, there is no reliability
- b. Concept of trustworthiness is core.

Validity

- a. Not an absolute, but based on the theoretical framework and data collection/analysis process.

Triangulation is advocated as the test for validity in qualitative research.

Typical key informant – information potential

Interview Subject	Information content – potential role in the evaluation[1 = qualitative 2 = quantitative]
Expert	<ul style="list-style-type: none"> • Theory of change[1→2 surveys] • Program antecedents [1] • History of and projected need for intervention [1] • Unique role for government vs other delivery options
Senior Manager	<ul style="list-style-type: none"> • Program origins and implementation [1] • Strategic management (program) issues (e.g., FPT relationships) [1] • Resource allocation (macro) [2] • Expected/actual results (macro) [2] • Alternatives (strategic/global)
Line Manager	<ul style="list-style-type: none"> • Project(s) origins and implementation [1] • Local management (project(s) issues (e.g., community/organizational relationships) [1] • Resource allocation at regional level (micro) [2] • Expected/actual results at regional (micro) [2] • Alternatives (program delivery)
Project Proponents (G&Cs)	<ul style="list-style-type: none"> • Project origins and implementation [1] • Local management (project) issues (e.g., community/organizational relationships) [1] • Resource allocation at project level (micro) [2] • Expected/actual results and project (micro) [2] • Alternatives (project implementation/delivery) [1]

Insert

..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\i
ndepth_e.pdf

Typical focus groups – information potential

Group type	Information content – potential role in the evaluation [1 = qualitative 2 = quantitative]
Client	<ul style="list-style-type: none">• Program implementation [1]• Program impact [1]• Field experiment [2]*
Management	<ul style="list-style-type: none">• Program implementation [1]• Program impact [1]
* Certain quantitative methods are ideally implemented in a small group setting. Conjoint analysis applied to program/policy design is an example that should be more widely used	

Focus groups are often seen as supplementary evidence designed to gather context about program implementation and impact, as well as ideas for program revision

The interaction among the participants means that the information whole is greater than the sum of the information parts.

Insert

F:\PROPOSAL\TEMPLATE\TECHNOTE\focusgroup_e.pdf

F:\PROPOSAL\TEMPLATE\TECHNOTE\fgfindings_e.pdf

Typical case studies – information potential

Case study type	Information content – potential role in the evaluation [1 = qualitative 2 = quantitative]
Maximum variation	<ul style="list-style-type: none"> Identify key patterns and variation (needs relatively large number.(>10)
Typical case	<ul style="list-style-type: none"> Identifies the norm [1,2]
Extreme (successes)	<ul style="list-style-type: none"> Best practices (feel good) [1]
Extreme (failures)	<ul style="list-style-type: none"> Corrective [1]
Politically critical	<ul style="list-style-type: none"> Gain wanted positive or suppress unwanted negative attention [1]
Convenience	<ul style="list-style-type: none"> Low cost – low information []

Analysing qualitative/quantitative data – mixed methods

Data Reduction

Quantitative

- Coding (pre-coding – post coding)
- Scales/indexes (Likert, magnitude)
- Factor/cluster analysis

Qualitative

- Coding (open/axial)
- Thematic development
- Typology/metaphor development

Quantitative and qualitative data **both** typically require some form of manipulation/processing

Data Analysis

Quantitative

- Univariate (one-way, summary...)
- Bivariate (cross-tab, correlation...)
- Multivariate (regression and other linear models)
- Instrumental Variables
- Structural equation

Qualitative

- Description
 - Summary
 - Thick
- Grounded theory
- Analytic induction
- Coding* (Categorization and connection)

Main differences

Quantitative research focuses on

- Measuring concepts (income inequality, cost-effectiveness, etc.)
- Establishing causality
- Generalizing from a sample to population
- Replicating and aggregating using standardized methods
- Discrete units of analysis

Qualitative research focuses on

- Explicating concepts and theories
- The actors' points of view
- Thick description
- Social processes



Eyewitness account of your birth

Your birthdate

Validity and reliability

- The commonly stated goal is to reduce bias and increase reliability
 - **Bias** is the difference between what is measured/observed and what is true
 - **Reliability** is generally defined as consistency in measurement



Reliable
Not Valid



Low Validity
Low Reliability



Not Reliable
Not Valid



Both Reliable
and Valid

<http://explorable.com/>

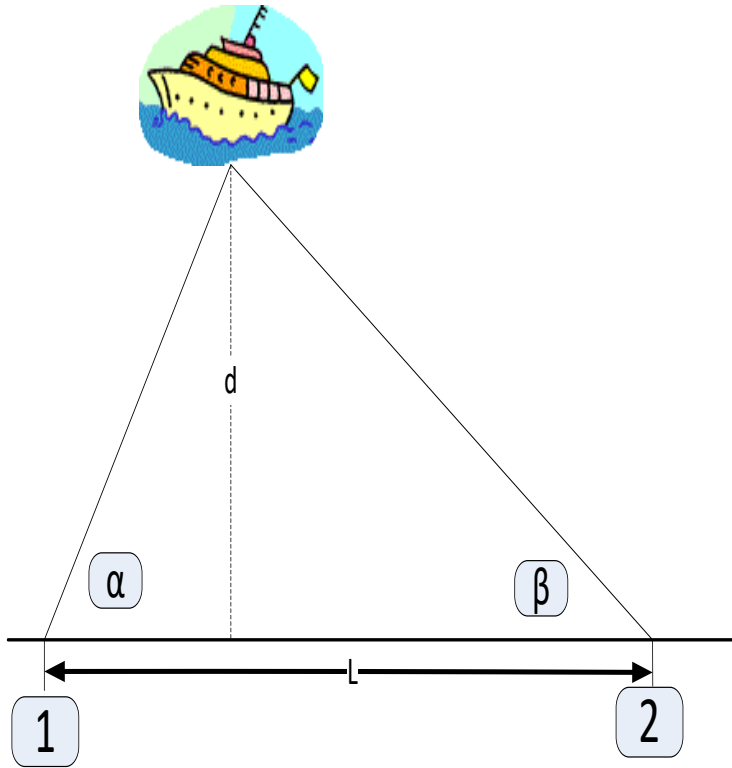
Interaction of quantitative and qualitative methods

- FIMCLA as originally intended is no longer relevant to farmers
- Expansion of FIMCLA into new loan uses (e.g., first time farm purchase) requires a needs assessment and risk analysis
- Lenders are finding the loan default and claims processing activities cumbersome and costly.
- The continued need for the program depends on several factors such as potential for interest rate increases, the need to finance recovery from crises (BSE, etc.).

The evidence that program use had effectively fallen to zero(because Canada had transitioned to a low interest environment) dominated the research process once it was revealed.

The single line of evidence of program decline cannot be “triangulated” with any qualitative evidence to modify the conclusion that the program was no longer relevant.

Triangulation – one more time



Two observers can “triangulate” the location of the boat (distance from the shore) by measuring the angles α and β and using the distance L and the law of sines.

Observer 1 and 2 measure the angles and the length L

**A man with one watch always knows the time.
A man with two watches is never sure.**

The key is that both observers are using the same theoretical framework (plane trigonometry)

Problems with triangulation

- Does not necessarily increase validity – competing perspectives fail to converge or collectively converge on a wrong idea
- May offer differing perspectives, but in social science this may not lead to less bias
- Mixing of quantitative and qualitative methods that draw from different theoretical frameworks usually results in the quantitative data dominating
- The analogy with surveying presents serious theoretical problems for mixing quantitative and qualitative methods

Using the problem to determine the distance of the boat from the shore, imagine observer 1 gave the angle as 23° while observer 2 gave the angle as “somewhat acute.” Further the helper hired to measure L decided to count the number of paces and not a standardized unit.

Triangulation has come to mean different things, depending on the ontological and epistemological framework of the researcher

Triangulation never starts after the data are collected. It starts with design and is integrated into all phases of study.

- Understand the nature of the evaluation questions and the information needed to answer each question
- Align the evidence that will address each question that
- Triangulate within a methodology :
 - Alternative statistical models
 - Simulations across several assumptions
 - Contrast the views of disparate key informants
 - Use multiple homogeneous focus groups to understand multidimensionality of experience and perceptions

By definition, triangulation cannot occur across methods

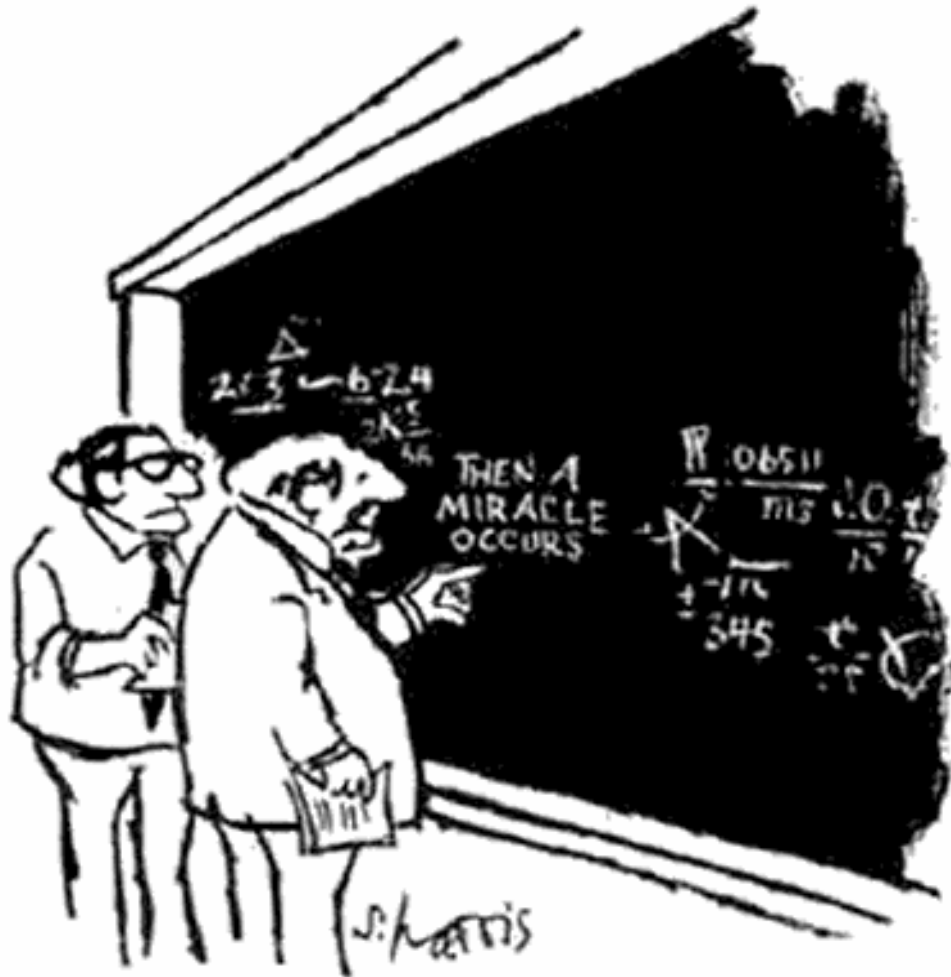
- Sample survey data on client satisfaction and key informant opinion about client satisfaction cannot be combined, any more than the time of day and the distance to Timbuktu can be combined.
- Key informants can voice legitimate perspectives on the content of the questionnaire, the nature of the sampling and what analysis might be appropriate
- Key informants perspectives and opinion on the results of a survey are also potentially valuable.

Multiple Triangulation

- **Data sources** – multiple sources (diverse key informants)
- **Investigators** – multiple interviewers
- **Multiple theoretical perspectives** – more than one theory of change
- **Multiple methodologies** – quantitative diversity and qualitative diversity

Within method (strategies within a method) vs. ***across method*** triangulation (dissimilar methods to assess the same unit)

“the flaws of one method are often the strengths of another, and by combining methods, observers can achieve the best of each, while overcoming their unique deficiencies” (Denzin, 1970a: 308).”



"I THINK YOU SHOULD BE MORE EXPLICIT
HERE IN STEP TWO."

© 1995 JEFFREY J. HARRIS

Distributed By Cotton-Engineering Ltd