PROGRAM INDICATOR SCREENING MATRIX (PRISM): A COMPOSITE SCORE FRAMEWORK

BONGS LAINJO
Cybermatic International, Canada

ABSTRACT

Over the last six decades, since many low and middle-income countries (LMICs) achieved independence, attempts of multi and bi lateral agencies to provide effective foreign assistance have been met with mixed results. It is in light of the above challenges that the framework Program Indicator Screening Matrix (PRISM) was designed. The model is defined by six criteria: specificity, reliability, sensitivity, simplicity, utility, and affordability. In an attempt to improve result-based management (RBM) in general and intervention data specifically, making foreign aid more focused and strategic with compelling evidence-based results, donors have increasingly teamed up and progressively introduced ubiquitous evaluation processes as an integral component of any program. A critical component that continues to confront many development aid stakeholders, is the ability to manage and improve the quality of data collected and reported. Current reviews of existing dynamics continue to be mixed. Hence a compelling need for a paradigm change cannot be adequately emphasized. Informed decisions are only as useful and constructive as the trustworthiness of the relevant data and quality. The general objective of this model is to strengthen the knowledge of implementing agencies and relevant stakeholders in mitigating indicator redundancies and optimizing quality results in program management. The model comprises a set of deterministic criteria simultaneously applied in identifying the most effective set of indicators in any thematic program area. It mitigates many program management nuisances by making indicators and data more reliable. A working group of experts use an analytical approach synonymous with the Delphi methodology. Each indicator during the assessment process is assigned a binary outcome (0,1) based on its performance with the relevant criterion. The resulting composite scores are evaluated against a bar established by these experts a priori. The model is thematically generic and inclusive.

Keywords: indicator-screening-matrix, results-based-management, gold standard, concordance, binary outcome, composite-score.

1 INTRODUCTION

The increasingly challenging and compelling demands for informed decision making have cumulatively contributed to the effective managing of data – data management. Further to this dimension is the copious amount of data required to better understand the dynamics of different programs. From a strategic framework perspective, the different levels of both the processes and results are spelled out very clearly in this paper. At the other end of the spectrum are myriads of historical data. The ubiquitous and evolving applications of predictive analytics have not only rendered decision making processes easier; they have also optimized the way we look at data. The propensity for large quantities of meaningful data is now the “modus operandi”.

In an attempt to improve intervention outcomes and make foreign aid more focused with evidence-based results, donors have increasingly teamed up and progressively introduced ubiquitous evaluation processes as an integral component of any program. In an attempt to streamline program implementation, donors in collaboration with recipient governments and other stakeholders have also promoted the availability and utilization of strategic frameworks (SFs). The current most frequently used SFs are the Logic framework and the Strategic Objective framework.

WIT Transactions on Ecology and the Environment, Vol 241,
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/SDP200111

A challenge that continues to confront many development aid Stakeholders, is the ability to establish an equitable, standard and inclusive strategy that includes, donors, national governments, implementing partners, program managers and beneficiaries. There is adequate evidence that the current approach of design and implementation strategies continue to show gaps and different levels of understanding of what programs are expected to produce.

This limited degree of understanding ultimately creates a gap between evaluators and key players in program implementation. This disconnect is so prevalent and detrimental that the level of "civility" between these two groups only exists between the presentation of the inception report and the draft copies of the evaluation. The generally brutal rebuttal given to evaluators after the first draft report is mind-boggling. Hence a tool like PRISM will serve and contribute to mitigating the sometimes, negative feelings between evaluators and key program players during a given evaluation process. Any program intervention focused on achieving meaningful and sustainable results will find this plausible and inclusive framework compelling and helpful. Its simplicity, comprehensiveness and thoroughness, cannot be adequately emphasized. And, if efforts by other model designers as confirmed by literature reviews are any indication, the introduction of more relevant and effective frameworks is a matter of time.

Over the years, funding agencies and national governments in developing countries have spent substantial amounts of money in developing and implementing programs. During the last decade, donors have been frequently faced with establishing sustainable and effective programs. This demand has also generated a compelling need for reliable and cost-effective results. These dynamics have been motivated by frequent demands from donor-country taxpayers for a more accountable and results-based programs.

In this paper, I have applied several decades of experience in data and program management – conceptualizing, designing and developing it, which combined, facilitates the mainstreaming of indicators PRISM. This stands for program, indicator, screening, matrix respectively. PRISM is defined by six criteria: Sensitivity, Reliability, Specificity, Simplicity, Utilization, and Affordability, with binary outcomes used in the screening process.

The available literature confirms that such models will effectively improve the dynamics of developing indicators in various thematic areas. The given circumstances, notwithstanding, the challenges ahead are familiar. As I learned in one of my early physics classes, all models are wrong, but some are helpful. The onus continues to be on the model designers and the relevant stakeholders to conceptualize, design, develop and articulate an inclusive thematic and potentially evidence-based model. PRISM represents such an initiative. Current and available frameworks do point at more and more positive outcomes.

Quality determinants are notions, procedures, and exercises, which add to the program's distinction. Quality indicators are rated using various criteria that take into account pragmatic, quantifiable, and dependable examination [1]. Many potential quality determinants and corresponding exist that reflect specific sections of the screening pathway. While coming up with quality indicators and determinants, it is expected that a portion of the essential parts of programmatic screening will be tended to indirectly rather than straightforwardly as described by Khampang et al. [2].

The PRISM model encompasses a number of deterministic criteria mutually applied to identify the most effective combination of indicators in the thematic program area. In the review of the information from various articles, one can see authors use various criteria for indicator screening such as specificity, reliability, sensitivity, simplicity, utility and affordability. The model consists of a set of deterministic criteria jointly realized in an effort to find the most effective collection of indicators in the thematic program section.

In the first article and according to Reiter et al. [1], individual criteria are used interactively with each other. They give an example of clear definitions as essential prerequisites for the reliability of valuable indicators, adding that high quality reliability of the indicator cannot be seen if the definitions have no clarity.

In a second article by Chan et al. [3], the criteria of sensitivity, reliability and simplicity is used to solicit the indicator set in a dual-tiered procedure of choosing, screening and rating of indicator. The results from the user-oriented degree of significance of the indicator suggest that users are sensitive to the details of landscaping, and consider the criteria of the parks' environmental simplicity relevant to management purposes. On the other hand, Reiter et al. [1] applies the criteria of affordability and specificity in the selection and screening of indicators. With the indicator feasibility, the co-authors test whether the cost of using the indicators was appropriate for an accurate validation of different qualities of healthcare.

In the third article, in which Khampang et al. [2] discuss the two approaches used – deductive and inductive to develop quality indicators with scientific evidence, essentially, refers to the criteria of sensitivity and utility. Khampang et al. [2] further endorse reliability as the degree to which the quality of the selected indicator is accorded consistent assessment.

In the fourth article, Scholte [4] uses the criteria of specificity (includes all) and states that indicators are of great essence and can be attained through a focus on the most predominant conditions. Consequently, the level of specificity increases the level of indicator usability to develop the quality of healthcare. Ultimately, according to Reiter et al. [1], the two greatest criteria used in assessing and screening the indicators are reliability and utility, which take place from two perspectives. The first perspective is the real utilization of the selected indicator outcomes for the decision-making procedure. The second perspective is the process of adding value in the context of the improvement of healthcare quality. The former covers the criterion of efficiency as the quality indicator in improving the experience of patients through the health care system. These findings confirm the appropriateness of the PRISM model.

The significant overlap among the criteria is also an indication of compelling experience and common ground. The articles by Chan et al. [3] and Khampang et al. [2] as examples, respectively, present a breakdown of most of the criteria (reliability, validity, feasibility, acceptability, being attributable, and sensitivity) presented in PRISM. Reiter et al. [1] address the "QUALIFY" model and under "scientific soundness" present criteria synonymous to those found in PRISM. According to Reiter et al. [1], individual criteria are used interactively with one another.

The author also uses the criterion of clarity to study and assess the definitions of the criteria, whether they are clear or ambiguous. According to the author, the criterion is a vital pre-requisite for finding a high level of reliability, specificity and sensitivity so, it is an essential criterion so needs attention [1]. Absence of clarity leads to randomness and insufficient interpretation. The author gives an example of clear definitions as essential prerequisites for the reliability of a valuable indicator, adding that a high quality reliability of the indicator cannot be seen if the definitions have no clarity.

The implications of these commonalities are significant and include the ability to replicate, compare and cross-validate models. The other implication is the degree of concordance among the different model authors. It serves as a pathway towards a likelihood of consensus with regard to the conceptual framework. While all these positive outcomes are reasons to be optimistic, the challenges remain fluid, compelling and undocumented. They also serve as topics for future thematic research. Some of them include:

1. A need to design and develop an optimum number of effective criteria;
2. A need to establish standardized definitions of all the criteria;
3. A need to develop a generic (thematic neutrality) number of criteria;
4. A need to develop an effective replicability strategy of the models;
5. Ongoing controversies notwithstanding, an attempt to operationalize randomized clinical trials (RCT) would also be helpful;
6. A need to digitize a potentially viable model; and
7. A need to promote the benefits that come with these models.

2 RESEARCH STUDY PROBLEM

The problem involved in this study is the mainstreaming of indicators. The PRISM model has shown that an excessive and redundant number of indicators is more productive in all these, in addition to an expensive data collection and analysis. According to Chan et al. [3], process indicators put more emphasis on the real care given to a patient, such as in diagnosis, treatment and communication. Indicators are theoretically efficient measures for gauging park conditions, while amplifying the consequences of many other conditions on justifiable management. The process of developing the indicators has to be well designed. To develop high quality indicators, there has to be a guideline-oriented methodology.

In models similar to the PRISM framework, both qualitative and quantitative indicators have to be test piloted before the real execution in the enterprise system [1]. An optimal selection of indicators will regularly contribute to the assessment plan, and is comprised of the assessment methods, analysis, and presentation of information. Optimally, selected indicators are an effective contributary pathway to the achievement of compelling and informed decision-making interventions. Consistent with the above highlights, PRISM is also aimed at:

- Demonstrating the effectiveness of conceptualizing, designing, developing and applying program indicator streamlining frameworks as illustrated by contemporary indicator mainstreaming models.
- Highlighting inter-relationships among similar and parallel models currently in use.
- Identifying the extent to which implementation replicability of existing models is possible with the focus on criteria that collectively delivers evidence-based strategies.

3 BACKGROUND

3.1 Logic framework and strategic objective

Program Indicator Screening Matrix (screening matrix) is the focus of this paper. The model, outlined later, is a pragmatic framework that effectively enhances program management operations. While different authors have developed a variety of program management strategic frameworks, the two most commonly used are the logic framework (log frame) and the strategic objective framework (SO). These two are both based on an apparent "causality" hierarchy. They also have the same process and results levels. From a log frame strategic framework perspective, (inputs, processes, output, outcome and goal/impact), the screening matrix is relevant and applicable at the output level, where implementing stakeholders are generally held accountable. The log frame structure is presented in Fig. 1.

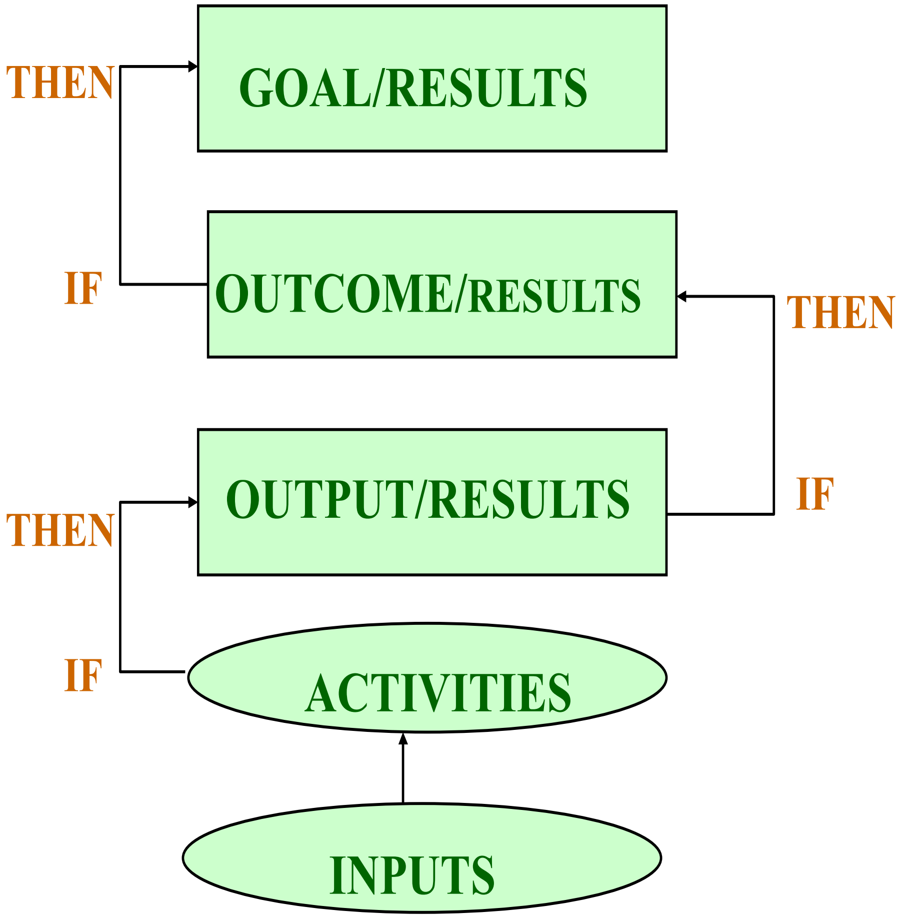


Figure 1: Log frame structure.

In Fig. 1, a health-care data management ecosystem is presented as an example of a more structured layout. The thematic framework (health) is also consistent with the log frame.

Discussions from the available literature indicate that two expansive methodologies are being utilized to create Quality Indicators (Q.I): a Deductive Methodology and an Inductive Methodology. A review led by Straus and Stelfox found that most of Q.I were created using a Deductive Methodology. This methodology proposes that the Quality indicators ought to be retrieved from logical evidence identified by significant concepts of quality-of-care. Further reviews recommend that key traits of good Q.I. are dependability, legitimacy, achievability, worthiness, being inferable, and affectability to change [3].

In Fig. 2, the model portrayed is an endeavor to map the underlying Program indicator with its corresponding higher-level results. With log-outline based structures, these levels include goals, output standards, and outcomes. As noted in Fig. 1, the critical target system will include various levels that relate intermediary results and develop critical achievements.

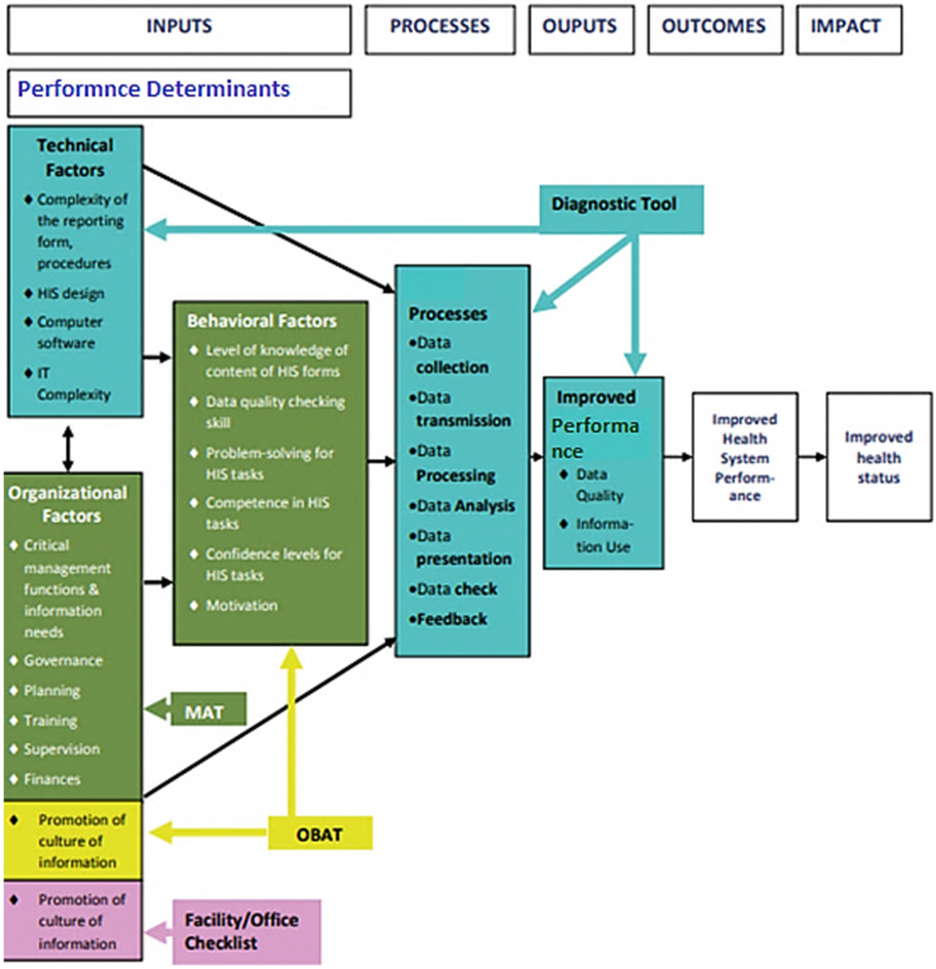


Figure 2: Screening matrix framework [5].

4 OVERLAPPING STREAMLINING CRITERIA AMONG SELECT MODELS

Based on the available literature, there is every indication that the urge to refine and improve program indicators is evolving and is progressively generating more interest globally. One challenge in operationalizing these models is the thematic inclusiveness and ability to replicate them. Other questions arise: Can one size fit all? Are there opportunities to adapt some models? How can the robustness of available models be established? Can the existing criteria be standardized to mitigate fluidity? Are there any "gold standards"? These questions and others need answers. Incorporating them into future research initiatives would be an easy task.

Table 1 is an illustration of how different criteria by different authors overlap. One implication is the implicit consensus that exists among these model architects. The other implication is from a triangulation perspective: the table does illustrate that there is compelling evidence to confirm the validity of these criteria.

Table 1: Indicator screening criteria distribution by author.

Model (i)	Author(s) and Year (ii)	Criteria (iii)	Theme(s) (iv)
PRISM	Lainjo [6]	Specificity, Sensitivity, Reliability, Utility, Simplicity, Affordability	Generic
QUALIFY	Reiter et al. [1]	Reliability, specificity, sensitivity, validity, risk, etc. (NB: Select Scientific Soundness category only)	Health
NA	Khampang et al. [2]	Sensitivity, Reliability, Utility	Health
NA	Chan et al. [3]	Sensitivity, reliability and simplicity	Urban Forestry
NA	Scholte [4]	Specificity and Utility	Physical Therapy
NA	Chan et al. [3]	Sensitivity, reliability and simplicity	Urban Forestry
PRISM	Simons et al. [7]	Risk, Availability and Appropriateness	Environment
NA	UK Studies	Sensitivity, reliability, validity, feasibility, acceptability	Generic
PRISM	Neta et al. [8]	Utility	Health
NA	Lip [9]	Risk, Specificity and Effectiveness	Health
NA	Lerch and Hermann [10]	Practicability, acceptability, sustainability and efficiency	Physical Therapy
PRISM	Leon et al. [11]	Efficiency and Utility	Generic
NA	Kwiatkowska et al. [12]	Affordability, efficiency	Generic
NA	Kopp et al. [13]	Utility and Availability	Clinical Health
NA	Grgic et al. [14]	Reliability, stability	Health
NA	Gidron [15]	Reliability and Validity	Health
PRISM	Aqil et al. [5]	Utility, Sensitivity	Health
PRISM	Ahmadi et al. [16]	Validity	Generic

5 THE PRISM MODEL

Standardized, well-defined and unambiguous criteria are a mandatory antecedent for the model to be effectively achievable. Pre-defined criteria contribute significantly to establishing a common ground and an understanding of the way forward. In the absence of a “gold standard” set of criteria, a research-based approach will serve (as demonstrated below) as a good basis for achieving a meaningful “standardization”. Such a strategy becomes even more compelling when dealing with groups of experts from different backgrounds and varied understanding of what each criterion may mean. Therefore, a consensus on a common and agreeable understanding of these criteria cannot be adequately emphasized.

5.1 PRISM methodology

The PRISM tool is a table aimed at extensively analyzing each indicator. This effort is executed by a team of experts, selected and grouped based on their relevant and appropriate expertise. An initial attempt is made to clearly describe the matrix, with its limitations and how it assists in addressing some of the challenges faced by program-implementing partners in establishing meaningful indicators. The final outcome of this exercise is a consensus or a degree of concordance (discordance) among the team members. See Figs 3 and 4.

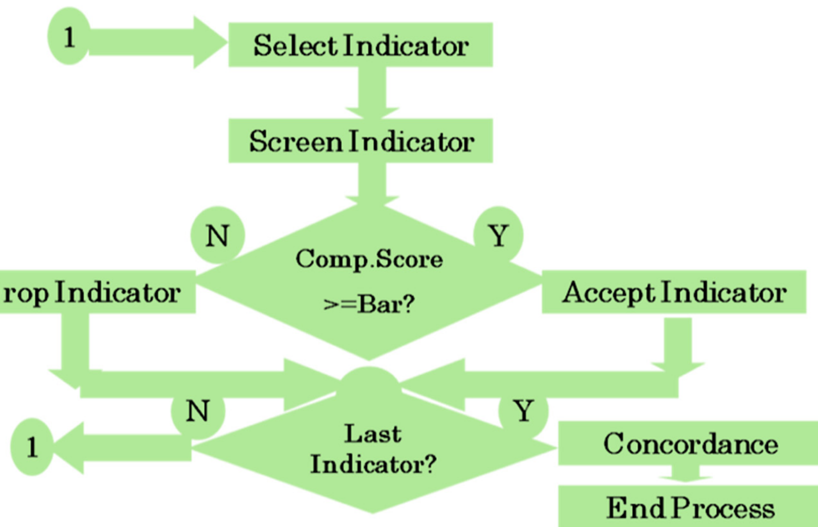


Figure 3: Graphic algorithm of the implementation process [6].

10 For Any Theme;
15 Are there any more themes?, NO →→ Step 70
20 Select thematic Indicator;
25 Has the last thematic Indicator been Screened? YES→→ Step 70
30 Screen thematic Indicator against Criterion; {Process continues through every Criterion};
35 Has Indicator been Screened against ALL criteria? YES>→→ Step 45
40 →→ Step 30;
45 Compute Indicator Composite Score {(Sum of "1"/ (Sum of ("1" + "0"))};
50 Compare Composite Score with pre-defined BAR;
55 IF Composite Score >= BAR, ACCEPT Indicator >>>>>>Step 20;
60 IF Not, DROP Indicator →→ Step 20;
70 Calculate Group Concordance;
80 END PROCESS

Figure 4: Procedure narrative of the implementation process [6].

5.2 PRISM general objectives

The objectives are two-fold: First, to strengthen the knowledge of Program implementing Agencies, Program Managers and other key stakeholders and emphasize a sustainable engagement in program management and implementation processes. Secondly, to address existing nuances, highlight the synergies that exist among the different result levels of the Strategic Frameworks and, hence, facilitate a common ground between potential evaluators and different interested parties.

5.3 PRISM specific objectives

The specific objectives are:

- To streamline monitoring plans by improving indicator causal links at all result levels;
- Mitigate the duplication of indicators;
- Establish authentic contributions between different result levels;
- Establish meaningful synergies among different result levels with an emphasis on: no lower level result can contribute to more than one upper-level result;
- Strengthen the program design;
- Promote a common understanding among key actors; and
- Minimize cost and optimize the number of indicators included in the program.

5.4 PRISM relevance

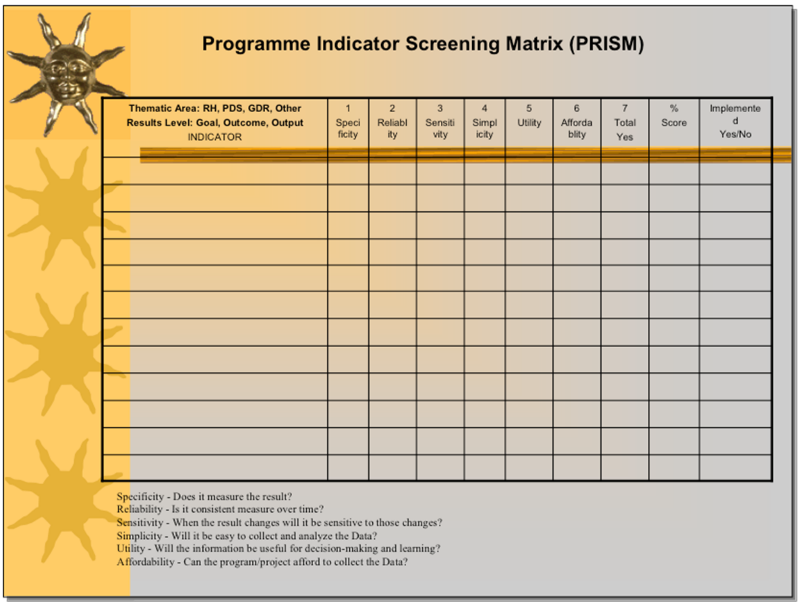
Relevant outcomes would include: improving intended and unintended intervention results and make funding more focused with evidence-based results; as well as establish more effective, continuous and sustainable synergies among frontline forces, Implementing Partners (IPs), Funding Agencies, Stakeholders and Beneficiaries.

5.5 Results and discussions

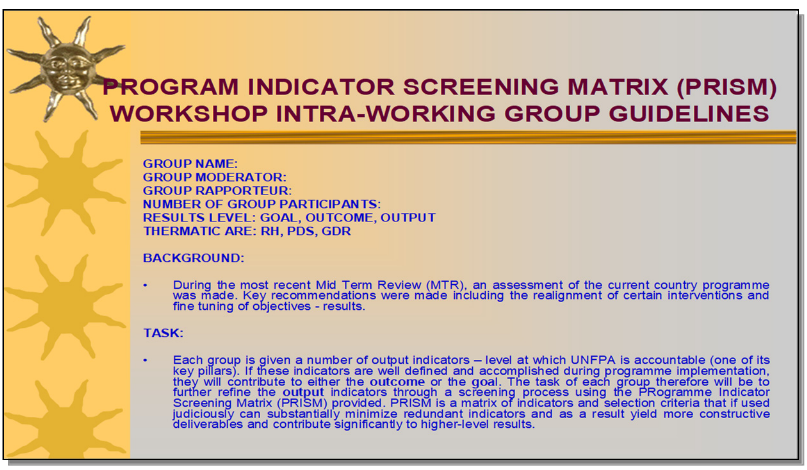
The definition of a team in this aspect is made up of groups and sub-groups with the latter serving as a sub-set of the former. As a rule, the two groups are composed of odd numbers. For instance, the team cannot work if only a single expert is available. On the other hand, the presence of two or more experts means that it is possible to establish one group – a coin can be tossed in case of a disagreement when establishing an indicator. Furthermore, if experts are available, a sub-group represents a group. That is, all the three members will work as a group, and the recommendations will be considered a group decision based on the degree of concordance. The process, in this case, is simple and obvious. Therefore, the majority decision (in this case, two out of three) prevails. This is how the rule of odd numbers applies, and the preceding description addresses outlier scenarios.

To the extent possible, this model works best if as many subgroups and groups can be established as possible without losing sight of the distribution (odd number of sub-group and group members). For example, if we have ten experts, we can easily create two sub-groups of five members each. In this case, the group will be ten while the sub-groups will be two. In general, the total number of sub-group members should be limited to eight. Experience has confirmed that, if there are more than eight members in a sub-group, some members become overwhelmed and tend not to participate fully. Finally, before each sub-group's work starts, the members are required to select a moderator and a rapporteur. The former then presents the sub-group findings during the final group meeting.

In Fig. 5(a), the table used to establish the number of acceptable indicators is made up of as many ROWS as there are indicators and TEN COLUMNS. The first row represents descriptions of each column. For example, in row one, column one, we fill in the relevant thematic area, result-level, and indicator. In the next eight columns (still on row one), we fill in the respective criteria to be used in screening the indicators. In the row below and subsequently, we have a table of binary elements, i.e. zeros and ones (0, 1). The former represents a corresponding indicator, which does not satisfy the criterion and the latter, a corresponding indicator that fulfills the criteria. The same process applies to all the criteria and corresponding indicators. Column 7 summarizes the scores in terms of the number of "yeses" (or 1s). The seventh column is the final score attained by each indicator, which is represented as a percentage of "yeses" in the row. The last column, the final outcome. tells us if, based on the scores (1s), we should go ahead and recommend the indicator or not. The "gold standard" for this exercise is 100%. That is the indicator that "yes" scores in the entire criterion qualify for implementation automatically. Criterion and description are described in Fig. 5(a). Self-descriptive Figs. 5(b), 5(c), 5(d), and 5(e). Detailed definition of each criterion is presented later.

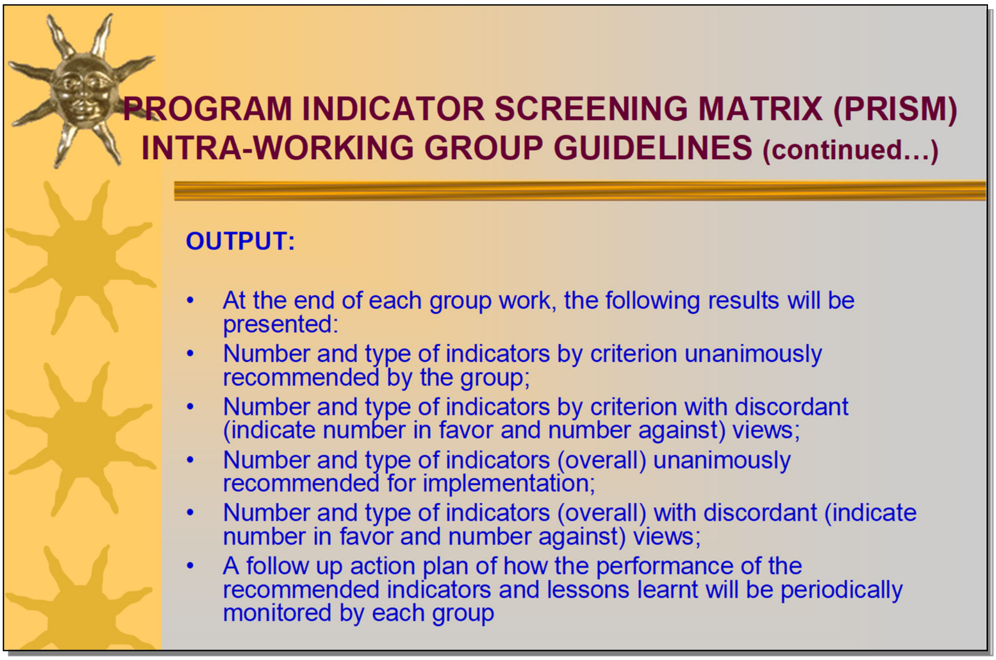


(a)

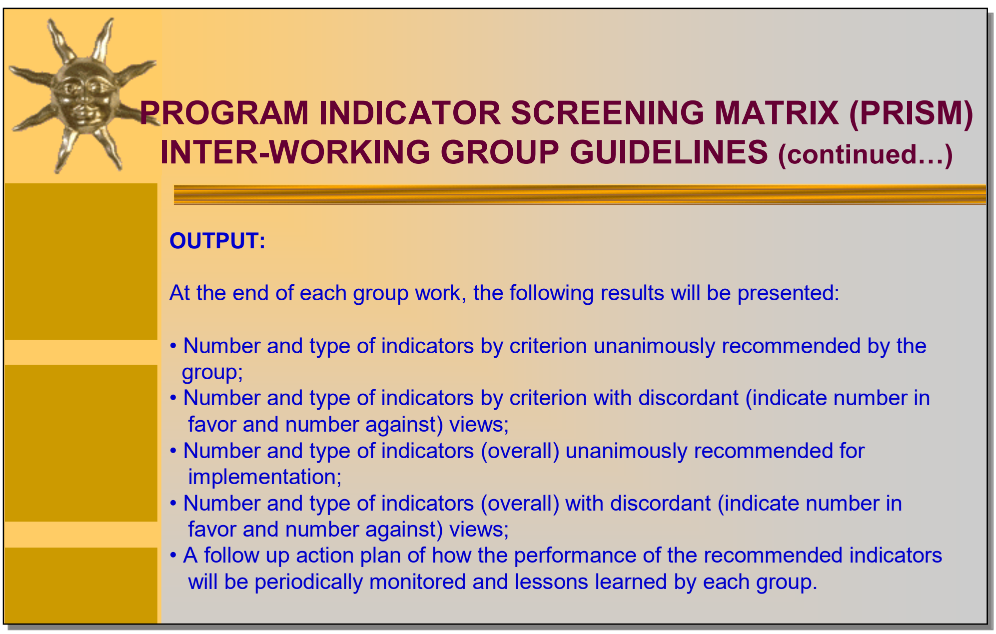


(b)

Figure 5: Program Indicator Screening Matrix (PRISM) case study procedures: (a) Part 1; (b) Part b; (c) Part c; (d) Part d; and (e) Part e. (*Source: Lainjo, 2013.*)

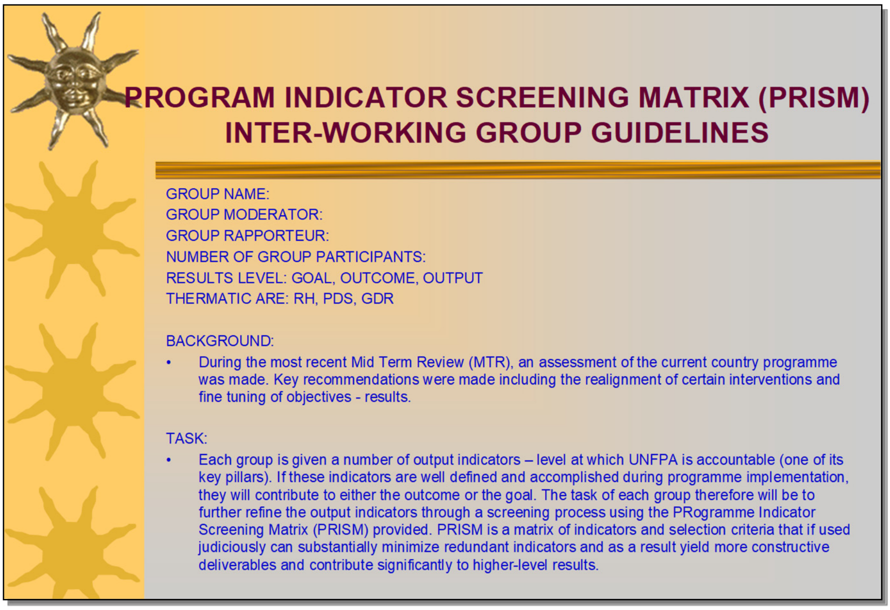


(c)



(d)

Figure 5: Continued.



(e)

Figure 5: Continued.

As this is a composite analysis, we need to remember that a final outcome is only valid when all the criteria are considered simultaneously. That is, the outcome identified in the last column. One can ask what happens if no indicator satisfies all these conditions? The answer is simple: before all the sub-groups begin their assignment, the team establishes an acceptable level a priori. For example, the team could agree before the exercise starts that any indicator that scores 70% (total "yeses" divided by sum of "yeses" and "nays") or decision level, will be considered acceptable. Sometimes, this bar can vary. For example, if the team recognizes that a certain threshold tends to admit too many redundant indicators, the bar can be raised higher in order to further refine and streamline the choices. The following paragraphs attempt to define some of the criteria as they apply to the matrix as well as explain research-based strategies that can be used to improve the quality indicators. A complete list of all the criteria is included in appendix A.

5.6 Sensitivity

Sensitivity is a test that tries to assess the stability of an indicator. For example, does the indicator continue to deliver the same result with a small variation of either the numerator or denominator? How does the result change when assumptions are modified? And does the indicator actually contribute to the next higher level? For example, an indicator at the output level accounting for one at the outcome level will yield a misleading result. If the same indicator accounts for two or more result levels simultaneously; it is not stable. As indicated earlier, any indicator that satisfies a criterion is given a "1" in the corresponding cell or a "0", otherwise. This assesses the indicator's stability as it highlights its ability to correctly come up with similar or closely similar results with a slight change of the primary variables.

5.7 Specificity

This refers to the likelihood of the indicator measuring the relevant result. In other words, is there a possibility that the result the indicator represents does not represent exactly what we are looking for? The specific characteristics of a test represent the ability of the test to appropriately correspond with the program's objective. Thus, to successfully improve the indicator's specificity, specificity needs to display a fixed characteristic of the test as well as represent a true negative rate. Specificity can be calculated as:

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True Negatives} + \text{False positives}}.$$

Based on this example, a difference may be seen in terms of fewer observations as compared to more observations. As a result, in comparison, the specificity needs more work than the sensitivity as the latter is already performing better. A large percentage of positive cases is indicated. Moreover, a high sensitivity test that produces negative results suggests the absence of a condition being measured. In contrast, specificity indicates negative test results. Improving specificity will thus involve a highly specific test that is efficient for the detection of a particular factor or element under study, if an individual tests positive. Similarly, it should not falsely indicate the presence of a factor that is absent. The specificity and sensitivity of a program depicted in a quantitative test depend on a cut-off value, which determines the limit between test results that are either positive or negative [9].

5.8 Reliability

This criterion is synonymous with replication that is, does the indicator consistently produce the same result when measured over a certain period of time? For example, if two or more people calculated this indicator independently, would they come up with the same result? If the answer is yes, then the indicator has satisfied that condition and hence a "1" is entered in that cell, or else "0" is entered. To improve the reliability of the indicator, it is significant to consider the different types of reliability.

5.8.1 Test-retest reliability

This test of reliability measures the reliability attained in the process of administering a single test twice to a group over a specified period. Then, the available scores from the different periods are correlated to evaluate the stability of the test over the period [17]. Thus, to improve reliability, the test-retest reliability offers the option of presenting a test twice to a group of individuals and establishing the degree of correlation.as a strong correlation coefficient would signify the score's stability.

5.8.2 Parallel forms reliability

This measures the reliability attained in the process of administering different adaptations of an assessment tool to the same group of people; however, both versions should contain items that explore the same knowledge base, skill, or construct, etc. To improve the reliability of the indicator, the two versions' scores are correlated to evaluate the result's consistency across them [18]. For instance, to assess the reliability of the indicator, a large set of items pertaining to the indicator is created. Then various variables are randomly split into two sets to represent the parallel forms.

5.8.3 Internal consistency reliability

This reliability evaluates the scale by which different indicators probing the same construct have identical results. Internal consistency reliability is further divided into "Average inter-item correlation" and "Split-half reliability" [19].

Average inter-item correlation is obtained by taking all variables on an indicator that explore the same factor, and determining the correlation coefficient for each variable as it takes the average of the correlation coefficients to yield the average correlation.

Split-half reliability involves the process of attaining a split-half reliability by first dividing in half all the indicators intended to explore the same factor. Two different sets of indicators are formed. In this process, the reliability is improved by administering the entire test to a group of people, computing the total score for every set and then obtaining the split-half reliability by determining the correlation between the two set scores [20].

6 FURTHER DISCUSSION: SCREENING MATRIX CONTEXT

The usage of the screening matrix tool gives an extensive image of the whole integrated system and the utilization of explicit tools could enlighten about a specific component. Nonetheless, the setting could make these findings unique. For instance, in a unified framework, the information accuracy and the utilization of data could be high, at an elevated level; yet the findings might show that the skills and knowledge to check the information data use quality at the lower level is restricted. One could state that the information quality checks or information usage be completed at a more significant level so consequently the framework would not have any shortcoming. It would be planned in such a way that lower level staff would be obliged to gather information and send it to the higher level.

In addition, as long as the functions were performed satisfactorily, then the framework would be operating at its optimal. Moreover, one could contend that if the senior administrators enabled staff at the lower level and enhanced their abilities, it would not just improve the staff capacity to utilize data for better service management, but it would likewise diminish supervision time and expenses. However, the interpretation to be conveyed depends on the senior administration, as it would have various ramifications for intervention and actions. Therefore, it is significant that each finding ought to be translated in line with the holistic structure and the setting in which the system works.

Another imperative to note is that we have given criteria for creating standards or provided benchmarks, which have become a regularizing standard for the platform. However, every setting needs to build up its own standards as there are situations where they would not be the same. Second, it infers that the standards level could likewise be under ongoing improvement. In this manner, we unequivocally propose that no standard or level is outright, but should be viewed as relative and utilized in the given circumstance accordingly.

7 FUTURE DEVELOPMENT

Practical encounters have demonstrated that in the future, refined tools can be utilized with singular criteria to make complex appraisal circumstances progressively justifiable so, therefore, can encourage a faster evaluation. For example, in the criterion capacity of measurable differentiation, the representation of the three-stage procedure could be improved after the first few encounters.

Unavoidable redundancies exist among singular criteria. In singular cases, it ought to be observed whether the criteria could be more plainly isolated from one another. For instance, the model's significance likewise captures aspects of usage, with the goal being that the criterion's benefit could be unmistakably centred on the indicator's practical benefit.

In the case of criterion reliability, it has been demonstrated that an improved database must be emphasized. In future, during the advancement of new indicators, an improved database should be made using explicit pilot tests (inter-rate at reliability and test-retest). The logical publication of the screening matrix tool and its pilot application would empower a debate, which could be the foundation for making further methodological headway with the tools.

With regard to future research, themes identified earlier remain valid and if potential research is conducted efficiently; there is every likelihood that indicator screening will continue to be a meaningful and evidence-based strategy for every potential stakeholder.

8 CONCLUSION

Screening matrix tools resemble any study questionnaire. However, in addition to utilizing interview procedures, the screening matrix employs observation, a review of information, perceptions, and testing. No specific skills are required to make use of a screening matrix tool, except for good observation and communication skills. Screening matrix tools are utilized when a comprehensive picture is required, and a connection between forms and performance needs to be determined. The challenges, notwithstanding, the model affirms that when embedded partakers are altogether associated with program development, a considerable level of synergism exists that encourages better organization establishment and promotes a typical comprehension of potential challenges in some programs in development zones.

APPENDIX A

Sensitivity

Sensitivity is a test that tries to assess the stability of an indicator. For example, does the indicator continue to deliver the same result with a small variation of either the numerator or denominator? How does the result change when assumptions are modified? And does the indicator actually contribute to the next higher level? For example, an indicator at the output level accounting for one at the outcome level will yield a misleading result. If the same indicator accounts for two or more result levels simultaneously; it is not stable. As indicated earlier, any indicator that satisfies a criterion is given a "1" in the corresponding cell or a "0", otherwise. This assesses the indicator's stability as it highlights its ability to correctly come up with similar or closely similar results with a slight change of the primary variables.

Specificity

This refers to the likelihood of the indicator measuring the relevant result. In other words, is there a possibility that the result the indicator represents does not represent exactly what we are looking for? The specific characteristics of a test represent the ability of the test to appropriately correspond with the program's objective. Thus, to successfully improve the indicator's specificity, specificity needs to display a fixed characteristic of the test as well as represent a true negative rate. Specificity can be calculated as:

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True Negatives} + \text{False positives}}.$$

Based on this example, a difference may be seen in terms of fewer observations as compared to more observations. As a result, in comparison, the specificity needs more work than the sensitivity as the latter is already performing better. A large percentage of positive cases is indicated. Moreover, a high sensitivity test that produces negative results suggests the absence of a condition being measured. In contrast, specificity indicates negative test results. Improving specificity will thus involve a highly specific test that is efficient for the detection of a particular factor or element under study, if an individual tests positive. Similarly, it should not falsely indicate the presence of a factor that is absent. The specificity and sensitivity of a program depicted in a quantitative test depend on a cut-off value, which determines the limit between test results that are either positive or negative [9].

Reliability

This criterion is synonymous with replication that is, does the indicator consistently produce the same result when measured over a certain period of time? For example, if two or more people calculated this indicator independently, would they come up with the same result? If the answer is yes, then the indicator has satisfied that condition and hence a “1” is entered in that cell, or else “0” is entered. To improve the reliability of the indicator, it is significant to consider the different types of reliability.

Test-retest reliability
This test of reliability measures the reliability attained in the process of administering a single test twice to a group over a specified period. Then, the available scores from the different periods are correlated to evaluate the stability of the test over the period [17]. Thus, to improve reliability, the test-retest reliability offers the option of presenting a test twice to a group of individuals and establishing the degree of correlation.as a strong correlation coefficient would signify the score’s stability.

Parallel forms reliability
This measures the reliability attained in the process of administering different adaptations of an assessment tool to the same group of people; however, both versions should contain items that explore the same knowledge base, skill, or construct, etc. To improve the reliability of the indicator, the two versions’ scores are correlated to evaluate the result’s consistency across them [19]. For instance, to assess the reliability of the indicator, a large set of items pertaining to the indicator is created. Then various variables are randomly split into two sets to represent the parallel forms.

Internal consistency reliability
This reliability evaluates the scale by which different indicators probing the same construct have identical results. Internal consistency reliability is further divided into “Average inter-item correlation” and “Split-half reliability” [19].

Average inter-item correlation is obtained by taking all variables on an indicator that explore the same factor, and determining the correlation coefficient for each variable as it takes the average of the correlation coefficients to yield the average correlation.

Split-half reliability involves the process of attaining a split-half reliability by first dividing in half all the indicators intended to explore the same factor. Two different sets of indicators are formed. In this process, the reliability is improved by administering the entire test to a group of people, computing the total score for every set and then obtaining the split-half reliability by determining the correlation between the two set scores [20].

www.witpress.com, ISSN 1743-3541 (on-line)

Validity

Validity refers to how well the indicator measures what it is meant to measure or how well it can reflect the reality. Improving the validity of the indicator increases the reliability of the test. Similarly, in this case, it is important to look at the various types of validity to determine the means of improving the overall outcome [21].

Face validity
This establishes that the indicator assesses the PRISM's construct. Thus, stakeholders can easily evaluate this kind of validity. Face validity is not entirely a precise form of validity; however, it is an essential component when enlisting stakeholders' motivation. If it cannot be proved that the indicator is an accurate assessment tool the task may face numerous disengagements. In improving the validity of the indicator, every single item needs to be related to the different types and components of the study to create the indicator appreciation measure. As well, there needs to be a reference to the indicator's movement for it to gain recognition and reflect the real assessment of the result or outcome [16].

Construct validity
It ensures that the indicator measures what the construct or the PRISM intends it to measure, and not any other variable. In this case, to improve the validity, there is a need to use a panel of professionals or experts familiar with the construct to assess the validity of the indicator. These professionals will be required to examine the applicability of the indicators and decide what they are intended to depict. By employing the skills and knowledge of the experts, the PRISM will be able to incorporate an indicator or indicators that adequately evaluate the proposed construct, rather than any other irrelevant element [22].

Criterion-related validity
It predicts current or future performance by correlating test results with another interest's criterion. To improve the validity of an indicator, it is important to associate the indicator with a standardized measure ability in the same scope. If the correlation between the new measure and the traditional measure is high enough, the indicator will gain great approval in regards to its assessment potential [23].

Formative validity
This assesses how efficiently a measure can provide information to improve the program being studied when it is applied. Thus, if the indicator can measure how well a particular factor is being processed, then the tool provides meaningful information that can be effectively used to improve the requirements of the program.

Sampling validity
This form of validity is quite similar to content validity as it ensures that the indicator covers an extensive range of areas within the concept or idea being studied. As not every item is included, the validity is improved by making sure that the different elements are sampled from all the relevant domains. Experts could effectively implement this process, which would help to ensure an effective exhaustion of all possibilities and that the content area is sufficiently sampled. Moreover, a good mix of experts would assist in limiting bias, as the assessment would adequately reflect the entire content area [24].

As Woodbridge et al. [24] emphasize, another way that validity can be improved in general is by making sure that the goals and objectives of the program are operationalized and clearly defined. Moreover, the expectations of the program and indicators should also be well outlined. As well, the assessment measures should be matched to the program's objectives and goals, and experts or professionals should review the entire process.

Simplicity

A convoluted indicator represents challenges at many levels. Hence, we must look for an indicator that is easy to collect, analyze and disseminate. Any indicator that satisfies these conditions automatically qualifies for inclusion. The zero/one process is then followed as indicated above.

The quality, reliability, validity, and acceptability of the indicators used to develop the results must not be enforced in a way that leads to complexity [25]. To improve simplicity, the indicators should ensure that the ethical aspects of research that may occur in several circumstances must observe traditional norms. Some may not regard this form of mundane science or inquiry to qualify as valid research. Nonetheless, these aspects are very significant to ensure simplicity and relevance of the research. Simplicity and other participatory research forms serve to bring appropriate stakeholders together in a flexible and reflective process that will maximize research possibilities and other success factors [26].

Utility

This refers to the degree to which information generated by this indicator will be used. The objective of this criterion is to assist in streamlining an indicator in an attempt to help the decision-making team make an informed decision. This can either be during the planning process or in the re-alignment process. The latter represents occasions when organizations are evaluating the status of their mandate. According to OECD [27], utility is defined as: "A summary term describing the value of a given data release as an analytical resource. This comprises the data's analytical completeness and its analytical validity. Disclosure control methods usually have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximize data utility whilst minimizing disclosure risk. In practice disclosure control decisions are a trade-off between utility and disclosure risk". In the PRISM model, utility and validity are used interchangeably.

Utility and the relevance of research have been a major topic requiring serious consideration for many years. The issue of utility appears to be more sensitive for programs and projects faced with wider competing demands and more severe limitations [15]. Further to this, there are divergent and multiple beneficiaries and users of the knowledge and the results generated from research such as service providers, action implementers, policymakers, and the wider communities. As well, the competitiveness of a nation is also linked to its innovation and research. Hence, it is highly important to develop research-based strategies that can improve the element of utility and the relevance of the research. The benefits associated with research may emerge from the research process rather than from the results or the final products. Thus, assessment of the utility and relevance of research becomes complex. In an attempt to improve the utility of the indicator, it is important to incorporate critical appraisal, intelligent choice, and the appropriate adaptation of various items in the application of the knowledge.

The problem's manifestation may be different in diverse programs, so a diversity of responses and information is required. Thus, it would also be important to incorporate a detailed analysis of problems that includes their extent and nature in specific areas of the program, and those concerning the program's goals and objectives and other underlying factors [10]. With regard to Information Technology, the necessary infrastructures such as management, personnel, facilities, infrastructure, and societal wealth affect the practicability of making use of the acquired knowledge, while cultural beliefs and attitudes can also influence the societal acceptability of the program. Consequently, the indicator applied in the program should address the above issues as one form of necessary research is the one that supports the formulation, implementation, and evaluation of policies.

Additionally, the indicator needs to link to the sustainable and efficient systems by presenting inclusive and comprehensive information, which forms the basis of the program's planning. Moreover, the indicator and the program must strike a balance between timely output within practical limits and the scientific severity. It is also important for the statistically and quantitatively proven results to be complemented by descriptive and qualitative case studies.

Affordability

This is simply a cost-effective perspective of the indicator in question. Can the program/project afford to collect and report on the indicator? In general, it takes at least two comparable indicators to establish a more efficient and cost-effective one. The one that qualifies is included at that criterion level. Then, the same process as outlined above is followed.

Kwiatkowska et al. [12] emphasize that as affordability is a cost-effective perspective of the indicator being measured, it evaluates the capacity or potential of the program to report and analyze the indicator. To assess the affordability of an indicator, there is a need for a detailed analysis of the efficiency of the indicators. In most cases, there is an evaluation of two or more indicators to identify the most preferable and profitable one, which can efficiently produce the intended goals and objectives of the program [28]. Similar to the above factors, affordability of an indicator is highly instrumental as it illustrates the most efficient and practical indicator. Indicators that are not cost effective are not considered in the program as they do not meet this criterion or respond to the above factors.

REFERENCES

[1] Reiter, A. et al., QUALIFY: Instrument for the assessment of quality indicators. *Dusseldorf: Bundes Geschafts Stelle Qualitats Sicherung*, 2007. www.researchgate.net/publication/267256474_QUALIFY_Instrument_for_the_Assessment_of_Quality_Indicators.

[2] Khampang, R., Teerawattananon, Y., Tantivess, S., Cluzeau, F., Foskett-Tharby, R. & Gill, P., Developing and testing quality indicators for the Thai Quality and Outcomes Framework. *Safety in Health*, **3**(14), 2017.

[3] Chan, C., HoiSi, F. & Marafa, L.M., Indicator development for sustainable urban park management in Hong Kong. *Urban Forestry & Urban Greening*, **31**, pp. 1–14, 2018.

[4] Scholte, M., It takes three to tango, developing and implementing quality indicators. for physical therapy: Lessons learned. Doctoral thesis, Radboud University Nijmegen, 2019.

[5] Aqil, A., Lippeveld, T. & Hozumi, D., Screening matrix framework: A paradigm shift for designing, strengthening and evaluating routine health information systems. *Health policy and planning*, **24**(3), pp. 217–228, 2009.

[6] Lainjo, B., *International Conference on Evaluation*, Toronto, Canada, 2013.

[7] Simons, L.E., Holley, A.L., Phelps, E. & Wilson, A.C., Screening matrix: A brief screening tool to identify risk in parents of youth with chronic pain. *Pain*, **160**(2), pp. 367–374, 2019.

[8] Neta, G. et al., A framework for enhancing the value of research for dissemination and implementation. *American Journal of Public Health*, **105**(1), pp. 49–57, 2015.
[9] Lip, G.Y., Assessing bleeding risk with the HAS-BLED score: Balancing simplicity, practicality, and predictive value in the bleeding-risk assessment. *Clinical Cardiology*, **38**(9), pp. 562, 2015.
[10] Lerch, J. & Hermann, B., Design your analysis: A case study on implementation reusability of data-flow functions. *Proceedings of the 4th ACM SIGPLAN International Workshop on State of the Art in Program Analysis*, pp. 26–30, 2015.
[11] Leon, N., Brady, L., Kwamie, A. & Daniels, K., Routine Health Information System (RHIS) interventions to improve health systems management. *Cochrane Database of Systematic Reviews*, **12**, 2015.
[12] Kwiatkowska, M., Norman, G. & Parker, D., Screening matrix: Probabilistic symbolic model checker. *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 200–204, 2002.
[13] Kopp, T. et al., Clinical improvement in psoriasis with specific targeting of interleukin-23. *Nature*, **521**(7551), p. 222, 2015.
[14] Grgic, J., Oppici, L., Mikulic, P., Bangsbo, J., Krustrup, P. & Pedisic, Z., Test-retest reliability of the yo-yo test: A systematic review. *Sports Medicine*, **49**, pp. 1–11, 2019.
[15] Gidron, Y., Reliability and validity. *Encyclopedia of Behavioral Medicine*, eds M.D. Gellman & J.R. Turner, Springer: New York, pp. 1643–1644, 2013.
[16] Ahmadi, Q. et al., SWOT analysis of program design and implementation: a case study on the reduction of maternal mortality in Afghanistan. *The International Journal of Health Planning and Management*, **31**(3), pp. 247–259, 2016.
[17] Neely, A., Adams, C., & Crowe, P. The performance prism in practice. *Measuring business excellence*, **5**(2), pp. 6–13, 2001.
[18] Velligan, D.I. et al., The reliability and validity of the MATRICS functional assessment battery. *Schizophrenia Bulletin*, **40**(5), pp. 1047–1052, 2014.
[19] Venable, J., Pries-Heje, J. & Baskerville, R., A comprehensive framework for evaluation in design science research. *International Conference on Design Science Research in Information Systems*, pp. 423–438, 2012.
[20] Danks, S. & Allen, J., Performance-based rubrics for measuring organizational strategy and program implementation. *Performance Improvement Quarterly*, **27**(1), pp. 33–49, 2014.
[21] Fornari, A. et al., Mentoring program design and implementation in new medical schools. *Medical Education Online*, **19**, 2014.
[22] Froyd, J.E., Borrego, M., Cutler, S., Henderson, C. & Prince, M.J., Estimates of use of research-based instructional strategies in core electrical or computer engineering courses. *IEEE Transactions on Education*, **56**(4), pp. 393–399, 2013.
[23] Subramanian, P., Ramkumar, N., Narendran, T.T. & Ganesh, K., PRISM: Priority based Simulated annealing for a closed loop supply chain network design problem. *Applied Soft Computing*, **13**(2), pp. 1121–1135, 2013.
[24] Woodbridge, M.W. et al., Implementation and sustainability of an evidence-based program lessons learned from the PRISM applied to first step to success. *Journal of Emotional and Behavioral Disorders*, **22**(2), pp. 95–106, 2014.
[25] Bonnefon, J.F., Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning*, **18**(3), pp. 379–393, 2012.

[26] Baskerville, R.L. & Wood-Harper, A.T., A critical perspective on action research as a method for information systems research. *Enacting Research Methods in Information Systems: Vol. 2*, eds L.P. Willcocks, C. Sauer & M.C. Lacity, Springer International Publishing, pp. 169–190, 2016.
[27] OECD, Glossary of Statistical Terms, 2005.
[28] Chan, A. & Chung, W., Teaching program design and implementation in a digital forensics context. *Proceedings of The 2015 NSF Workshop on Curricular Development for Computing in Context*, pp. 4, 2015.