Reviewer's report

Title: Advancing the application of systems thinking in health: A realist evaluation of a capacity building programme for district managers

Version: 3 Date: 8 May 2014

Reviewer: Jens Byskov

Reviewer's report:

The comments in my first review have been addressed.

What remains can be classified as discreitional revisions for my additional comments on realist evaluations (RE) as it is presented.

I am not clear enough on all terminology and definitions within the methodology, but see a range of options for the scope, methods and tools for RE's. The specific case and situations to me become a setting for testing, weighing and discussing several of these options. In the process the case results do not stand out as very concrete for the future shaping of the program in question where choices still remain for what to give highest priority in the further shaping of the evaluated program. This is fine if it is clear that this is in fact a largely methodological paper that illustrates a quite full range of RE approaches. Some of the methods and the scope are a bit of an overkill, but this is of course a result of its own. If agreed this may need to be considered in the final version.

The above comment is also due to and observation of to me not fully consistent terminology in the various introductory and discussion sections and graphs. Internally the graphs paint a coherent picture, but I am not sure that I fully understand or agree with the face value of all the boxes and arrow links and directions.

It is not all methods and tools that come out with results that are giving a convincing picture or are not open to more than one interpretation. However in RE it is the combination of them all that approach a realistic actionable result. Still looking at the quantitative results in e.g. the outlier Taluka choice and comparison they seem to in isolation violate considerations of quantitative result validity. Are we here caught between the two main qualitative or quantitative research paradigms? To me RE is really a more qualitative research approach and results are valid for the specific programs or settings in their own context in showing important processes toward achieving outcomes. Whether such outcomes materialize is less feasible to prove as the whole context and time factors can never be completely encompassed.

For example I am looking for some kind of hierarchy in context-environment-sociocultural factors. I am not completely sure what includes what.

That people make the changes is true as they implement the program and follow or supplement processes, so they are not working in an open space situation. It
is also important to know what is meant by people. I would rather say that it is the
decided processes and the people involved in each of them that explain results.
Depending on the guiding processes the people involved at each stage involves
politicians, managers, health workers, users and the public at large. Therefore it
is not clear everywhere which type of people that are involved.

Do also check that the use of terms process and function is consistent.

Section 1.4 is fine, but I wonder if it could be split up and merged into the
previous sections.

The nature of middle range theory could briefly be further explained in order to
understand what is not guiding the paper.

Some places it is not completely clear if it is the Taluka or the district of same
name which is meant.

The use of the word efficacy seems to differ from the conventional use of
effectiveness in ideal controlled (context independent) situations. As used it looks
more like a person level efficiency.

I do not need to see the paper again before publication, which I recommend.

Jens Byskov

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

The comments in my first review have been addressed.

What remains can be classified as discrentional revisions for my additional
comments on realist evaluations (RE) as presented.

I am not clear enough on all terminology and definitions within the methodology,
but see a range of options for the scope, methods and tools for RE’s. The
specific case and situations to me become a setting for testing, weighing and
discussing several of these options. In the process the case results do not stand
out as very concrete for the future shaping of the program in question where
choices still remain for what to give highest priority in the further shaping of the
evaluated program. This is fine if it is clear that this is in fact a largely
methodological paper that illustrates a quite full range of RE approaches. Some
of the methods and the scope are a bit of an overkill, but this is of course a result
of its own. If agreed this may need to be considered in the final version.

The above comment is also due to and observation of to me not fully consistent
terminology in the various introductory and discussion sections and graphs.
Internally the graphs paint a coherent picture, but I am not sure that I fully
understand or agree with the face value of all the boxes and arrow links and
directions.

It is not all methods and tools that come out with results that are giving a convincing picture or are not open to more than one interpretation. However in RE it is the combination of them all that approach a realistic actionable result. Still looking at the quantitative results in e.g. the outlier Taluka choice and comparison they seem to in isolation violate considerations of quantitative result validity. Are we here caught between the two main qualitative or quantitative research paradigms? To me RE is really a more qualitative research approach and results are valid for the specific programs or settings in their own context in showing important processes toward achieving outcomes. Whether such outcomes materialize is less feasible to prove as the whole context and time factors can never be completely encompassed.

For example I am looking for some kind of hierarchy in context-environment-sociocultural factors. I am not completely sure what includes what.

That people make the changes is true as they implement the program and follow or supplement processes, so they are not working in an open space situation. It is also important to know what is meant by people. I would rather say that it is the decided processes and the people involved in each of them that explain results. Depending on the guiding processes the people involved at each stage involves politicians, managers, health workers, users and the public at large. Therefore it is not clear everywhere which type of people that are involved.

Do also check that the use of terms process and function is consistent.

Section 1.4 is fine, but I wonder if it could be split up and merged into the previous sections.

The nature of middle range theory could briefly be further explained in order to understand what is not guiding the paper.

Some places it is not completely clear if it is the Taluka or the district of same name which is meant.

The use of the word efficacy seems to differ from the conventional use of effectiveness in ideal controlled (context independent) situations. As used it looks more like a person level efficiency.

I do not need to see the paper again before publication, which I recommend.

Jens Byskov

I have no competing interests in any of the areas mentioned in the HARPS web form