# Rejoinder

The authors have performed a systematic literature review of cumulative voting (CV) and CV analysis methods. In addition, they propose an analysis method that identifies prioritization items with equal priority. The paper is well written, and the topic is relevant for IST Journal. There are, however, a number is questions that need to be addressed before being considered for acceptance.

-     The title does not match with the content, as the SLR is supposed to be the main emphasis (but not mentioned in the title).

Answer:

Title changed to: Equality in Cumulative Voting: A Systematic Review and Improvement Proposal

The SLR is one contribution, but that is just the means to be able to then empirically validate our approach, i.e. ECV.

-     The objectives of the paper need to be stated more clearly. The authors say they want to aid decision-making, without mentioning the actual decisions. Is the assumption that decision-making about methods and techniques is done independent of the size of the problems?

Answer:

Actual decisions are described in the first paragraph of the first section of the paper. We’ve also added two examples now

The sentence “Is the assumption that decision-making about methods and techniques is done independent of the size of the problems?” is not clear to us. Could we get a clarification in this matter?

-     If this is a systematic literature review, then the results of the study selection should be made available.

Answer:

Selected studies are listed in Appendix A.

- Study selection criteria need to be made explicit.

Answer:

All study selection criteria are described in Section 5.2 Study selection and Tables 3 and 4. Please inform us of any additions/changes you believe are necessary to that data.

Also: Is study classification done by one researcher (and quality evaluation done by another one) sufficient?

Answer:

Part of study selection and data extraction was done by two researchers and the results were evaluated using inter-rater agreement analysis. The analysis showed sufficient level of agreement.

-     If the scoping of the SLR is on software engineering (not sure this has been said somewhere), why papers from forestry and government selection were considered as well? Are these the only empirical studies from outside software engineering?

Answer:

The short answer is: Yes. ☺

Studies from outside software engineering were included to increase the amount of data for the evaluation of ECV. The distribution of studies among different fields is shown in Figure 7: Distribution of studies over time.

-     It does not become clear how the SLR and the ECV approach are related to each other. Is ECV somehow suggested/motivated by the SLR?

Answer:

One of the goals of the SLR was to reveal any alternatives to ECV. SLR did not reveal any such alternatives and we thus developed and then validated ECV against data we found through the SLR (by e.g. contacting authors).

The need for ECV is described in more detail in Section 6.

-     The flow of presentation is a bit difficult to follow. The SLR design is presented in Section 5, the results are presented in Section 7, threats to validity are discussed in Section 8.4. A more cohesive presentation of the content would facilitate better understanding  
- The related work section is very short, actually discussing just one paper. This is insufficient for this purpose.

Answer:

We have now instead strengthened the background section (Sect 2). We are uncertain if changing the complete flow of the paper is necessary since other reviewers did not comment on this as far as we can tell? If the editor feels this is necessary we’d be happy to comply, what we have done for now is to clarify the presentation (end of Sect. 1) and added small parts all over the paper to remind the reader why we present our results later (when we also validate ECV).

- There are a number of specific questions and recommendations related to compositional data analysis as raised by reviewer 2. Overall, the paper looks interesting and promising. The authors are requested to prepare a revised version in which the comments and suggestions are taken into account.  
  
Reviewer comments:  
  
Reviewer #1:  
  
- There are a few minor problems that should be resolved before publishing the paper. On page 6 feminine form is used instead of infinitive (She may manipulate... to her but not to the other...). The same problem is on page 20 lines 526 to 531 (she and her). On page 19 line 466 - studyy -> study. Also the description on this page is very general. Maybe it could be replaced by a reference describing the theory of quality evaluation.

Answer:

We’ve fixed the minor things above. In addition, the goal of the description was to justify the selection of the quality evaluation criteria. Quality evaluation is an important part of the paper thus we would like to keep this part if possible.

- On page 20 line 518 authors stated that sometimes no results is more interesting than a significant result. You should provide more evidence and appropriate reference to the statement or reformulate the statement.

Changed to:

“Additionally, sometimes no result is more interesting than a significant result, i.e. it may reveal important gaps in existing knowledge.”

- On page 21 line 572 "for us" could be omitted - "In order to perform a more..." On page 22 Table 5 - The rating should not be black and white. There could be an academia study that has high reality level. Maybe it is better to use more values.

Answer:

All minor fixes done.

The rating is not black and white, three aspects are used to rate study realism: subjects, setting and scale. We believe that these three aspects, and their different levels, can be used to catch these cases. We would be very interesting to hear if there might be a case that cannot be covered by these three aspects.

- Page 24 lines 604-607 It is not clear why the equal requirements could be interchanged between releases. There might be two very important requirements rated as equal, however both must be in the same release. Please re-evaluate this finding and try out different real-world examples to verify the correctness of this statement.

Answer:

We clarified the statement in the paper: “If two or more requirements are considered equal they can be interchanged between the releases regardless of their priority. That allows other criteria, such as cost or effort, to be used as sole indicators for planning that particular release.”

- On page 34 line 617 the sentence should be "a significant difference in the tests,..."

Answer:

Changed as suggested

- On page 25 equation 6: It is not clear what is the meaning of the \* sign. References 36 and 38 are not clear. You should provide additional information regarding these references or remove the references.

Answer:

Added description: „"\*" denotes all rows of corresponding column.”

Added URLs to the references:

[45] K. Rinkevics, R. Torkar, Data extraction and quality evaluation results (2011). URL http://rinkevic.wordpress.com/2011/11/26/data-extraction-and-quality-evaluation-results/

[47] K. Rinkevics, R. Torkar, ECV implementation source code in R (2011). URL http://rinkevic.wordpress.com/2011/08/14/ecv-implementation-in-r/

## Reviewer #2:

The manuscript deals, at least, with three goals: (a) extensive review of CV prioritization methods; (b) state the compositional character of CV data; (c) develop testing of equal priorities using compositional methods. These goals are not specified by the authors in this way (lines 30-36) but they come from my understanding of the manuscript.  
  
Goal (a) is fully accomplished; goals (b) and (c) are partially accomplished although further clarification and research is necessary. In general, the manuscript is well written and has a considerable potential impact if some points are improved. I think that the manuscript can be accepted for publication after a major revision.  
  
Important points to be changed/improved.  
  
A1.- Goals for the manuscript should be clarified (lines 30-36). If  
the only goals are (a) and (c), the discussion on the applicability of  
compositional data (CoDa) analysis to CV may relay in the only  
reference [10], where the authors state the compositional character of  
CV data. However, there is a extensive discussion on compositional  
data methods in the present manuscript thus pointing out that goal (b)  
is implicitly assumed. Goal (b) should be thought separately from (c)  
because the potential application of CoDa analysis to CV largely  
exceeds testing of equality of two prioritization items.

Answer:

Added the following:

“CV results correspond to special type of data -- compositional data. Principles of compositional data analysis are described in this paper.”

A2.- It is not clear what is the authors' opinion about the  
compositional character of CV data. On one hand, they tell the reader  
the many problems of CoDa analysis (section 2.4, lines from 280 to the  
end of the section; section 7.2.1). On the other hand, the authors  
seem to support the compositional character of CV data and log-ratio  
analysis and, if fact, they use these methods to propose ECV. I  
understand that what the authors mean when speaking about problems of  
CoDa analysis refer to the pitfalls when CoDa are treated not taking  
into account their compositional character. However, it is explained  
in a confusing way.

Answer:

|  |  |
| --- | --- |
| Original | New version |
| “Compositional data analysis has, however, serious limitations.” | “When doing analysis of compositional data one must take into account that compositional is a special type of data and should be analysed differently than other data types” |
| Problems with Compositional Data Analysis in Primary Studies | Problems with Data Analysis in Primary Studies |
| A few primary studies, as revealed by the systematic review, have problems with the analysis of compositional data. | A few primary studies, as revealed by the systematic review, have problems with the data analysis.  These studies disregard the compositional nature of CV results. |

A3.- The manuscript describes important features of CV data that  
characterize them as compositional. This is clear in (section 2.4,  
lines 265-272). This description corresponds to the principle of  
"scale invariance" stated by Aitchison (1986) and then reformulated  
several times. In order to clearly state the compositional character  
of CV data, the principle of subcompositional coherence should also be  
addressed. The manuscript describes features of CV that perfectly  
match this principle (for instance, the hierarchical structure.

Answer:

Added the following text:

“Another property of compositional data items is subcompositional coherence. Consider that two compositions are analysed. One composition is a subcomposition of the other. Subcompositional coherence means that the results of the analysis are the same for the common parts of the compositions. This property is important for the analysis of HCV results. Statements that are made regarding each smaller group of prioritization items are also true for all items prioritized with HCV.”

A4. Method ECV is a traditional hypothesis testing on the log-ratio of  
two priorization items, with the null hypothesis that the log-ratio is  
null. The fact that a log-ratio can be considered as one component of  
an ilr-transformation does not justify the introduction of  
ilr-coordinates as it is done in section 2.4.2. The analysis of  
log-ratios and log-contrasts was developed by J. Aitchison, see  
Aitchison 1986, about 20 years before the introduction of ilr. Section  
2.4.2 introduce a particular ilr-transformation that is not used  
latter on in the manuscript. Therefore, section 2.4.2 appears quite  
obscure to a standard reader. I would suggest to introduce  
ilr-transformations based on a sequential binary partition (SBP)  
because it perfectly fits the hierarchical structure of HCV data and  
many consequences may be inferred. If the authors prefer leaving this  
discussion for further papers, I would recommend to suppress section  
2.4.2 and just mention that ilr-transformation is presently a key tool in CoDa analysis.

Answer:

Removed the formula for particular irl transformation.

A5.- ECV is presented as a method to decide about equality of  
priorization of two items under CV. In section 6.2 the  
non-transitivity of equality is well described, but it is not clear  
how the two criteria of grouping equal items are used. This subsection  
should be developed. Suggestion: test equality of two groups of items  
using the balance (log-contrast or ilr-coordinate obtained through a  
SBP). Assume (x\_1, x\_2) and have been declared equal. Then sqrt(2/3)  
log ( (x\_1 x\_2)^{1/2} / x\_3)  , a balance between the group (x\_1,x\_2)  
and the single component group (x\_3), can be useful two test equality  
of x\_3. This can be generalized to any size of the groups.

Answer:

Added an explanation:

“Current implementation of ECV (available from []) does not include the division of items into groups. In this study division is done manually, so that each two items in a group are equal.”

We would like to include the grouping of the prioritization items in future work.

A6.- Section 7.2 reviews a some techniques commonly used to present CV  
results. Most of them ignore the principles of compositional analysis  
and consequently they can be misleading. Authors do not use these  
procedures but they do not criticize them. In sub-section 7.2.1  
criticism seems to be assigned to the fact that CV-data have  
compositional character (see point A2.-). The problem is not that  
CV-data is compositional or the compositional methods themselves but  
the use of methods designed for real variables when applied to CoDa.

See A2

A7.- Line 973-975. Raw CV data cannot be normal distributed because  
its support is limited, as mentioned by the authors. However,  
log-ratios can be normal distributed (not necessarily). The first  
assertion cannot justify the use of non-parametric methods to  
log-ratios as those used in ECV. Non-parametric methods can be used  
anyway but this is not supported on the non-normality of raw CV-data.

Answer:

We have now clarified how the test that was performed:

„In our case, the CV result data obtained from the primary studies identified by the systematic review, were tested for normality using the Anderson-Darling test. Before applying the test the data was transformed using methods of compositional data analysis. To compute the test we used the method adtestWrapper from R language library robCompositions.”

## Points to be amended

B1.- Authors use strong statements when describing situations that are  
likely to be optative. Some of these statements should be relaxed.  
Examples:  
line 239: "These values must be summed together to form the final  
priority..." ("must" is too strong).  
line 633: "non-parametric estimation of the distribution function is  
needed." ("needed" is too strong; there are other alternatives:  
parametric, simulated-MonteCarlo, bootstrap, Bayesian...)  
line 684: "When equal items are determined they must be divided into  
groups of equal items" (again, "must" is too strong, "can be" seems  
better)  
line 803: "compositional data must first be transformed using  
isometric log-ratio transformation..." ("must"; this is not true,  
there are non-isometric valid alternatives; "is convenient" may be a  
suitable change in this statement).  
Please, check other statements of this kind throughout the manuscript.  
They reflect some dogmatism not based on any evidence.  
  
  
B2.- line 466: studyy (a typo)  
  
B3.- In most mathematical contexts equations must be punctuated and  
they end with a comma, a semi-colon or a stop. Some equations in the  
manuscript are punctuated, but most of them are not. I would suggest  
to punctuate all equations unless the editing guidelines of the  
journal specify the contrary.  
  
B4.- There are some sentences starting with a numeric reference (e.g.  
line 502: "...study realism. [32] suggests...") This kind of sentences  
should be changed to avoid starting with a reference number. Authors  
may be cited. Or "Reference [32] suggests...", etc.  
  
B5.- Lines 643-653: there is a confusion on notation ilr(k,l). The  
indexes k,l denote columns but in eq. (7) they are values of CV. Also  
"ratios" is really referred to "log-ratios". I think that "log-ratio"  
avoids confusion.

B6.- Check all references. For instance, in [21] some accents are  
missing. References [36] and [38] are not complete, they do not allow to access them. There are no final pages in references [35] and [52].

Answer:

Changed ratio to log-ratio

Clarified the symbols, changed equation (7) an.d replaced ilr(k,l) with ci

Notes B1 to B6 amended as proposed by the reviewer.

Suggestions:   
C1.- The presentation of the ilr-transformation is not justified in  
the manuscript in its present version because ECV only uses simple  
log-ratios as log(x\_i/x\_j). This may be supported on the log-ratio  
analysis as presented in Aitchison (1986) book. However,  
ilr-transformation based on sequential binary partitions (SBP) can be  
the basis of a further research of hierarchical CV-data (without  
common items on the groups). Further research with common items in  
different groups of the hierarchy should be studied using  
non-orthogonal (non-isometric) coordinates. I would suggest the  
introduction of ilr based on SBP better than the particular case  
described in sub-section 2.4.2. I would present SBP-based ilr as a  
tool available in CoDa analysis that will be useful in further  
developments. It will be particularly interesting for compensation of  
priorizations when using HCV. Note that standard priorizations in HCV  
are not subcompositionally coherent and that the authors are near to  
give a compositional alternative.

Answer:

If possible we would like to add the above suggestion to the Future Work section.

C2.- The problem of zero-priorization is here treated as a problem of  
"under detection limit", or "rounded zero" problem. Consequently, the  
imputation technique used is that introduced by [21] J.  
Martín-Fernández, C. Barceló-Vidal, V. Pawlowsky-Glahn (2003).  
However, there is an alternative that authors may have into account. A  
brief description follows:  
The designer of the survey cannot be considered as an stakeholder.  
However, stakeholders implicitly recognize the designer by consensus.  
Then, opinion of the designer has some weight on items present in the  
survey. This weight is uniformly distributed on the items because the  
designer does not priorize some items relative to other ones. If there  
are k items and the weight of the designer is w (integer or not), then  
each item has a prior CV value of w/k to be added to stakeholder  
opinions over a total of 100+k (w/k). This approach does not respect  
the ratios between items assessed by stakeholders. But one can assume  
that stakeholders are only assessing ratios roughly: in fact, when a  
stakeholder assigns 0 to an item, this is not relative to anything  
thus showing he/she is not accurate with relative values.

Answer:

Interesting but doesn’t this potentially seriously limit the usefulness of our approach?

If there are two items to rank: Even when a stakeholder assigns 0 to one item, we would claim that they make an explicit or implicit assumption, i.e. the item is valued 0 compared to item #2 which is valued, for example, 50. 0 can mean “completely irrelevant” in this case (implicit), but it can also mean that it is irrelevant only against item #2 (explicit). Thus, it feels intuitively that a rounded zero approach is the right way to treat this. However, this could of course be evaluated in an empirical study…

C3.- Once the compositional character of CV data is accepted, there is  
a way of making easier the task of stakeholders. Just leave the  
stakeholder to assign as many points as he/she wants to the items  
following his/her opinion. At the end, the assigned points add to a  
total. The compositional methods (log-ratios, log-contrasts, balances,  
ilr-coordinates, etc.) remove this total whichever it is. Therefore,  
the constraint of summing 100 or 1000 is completely irrelevant and can  
change from one stakeholder to another one. In HCV, this means that  
removing hierarchy should be done using compositional methods, which  
up to my knowledge have not been described before.

Answer:

If sum constraint is removed then it would no longer be CV? One could study how the presence of a sum constraint affects the prioritization. But this discussion is out of the scope of the paper as the reviewer indicates ☺

C4. Again assuming the compositional character of CV data, weighting  
stakeholders should be done using powering (the operation in the  
simplex) better than using a multiplication that will be removed in a  
log-ratio.

Answer:

A very interesting suggestion. We would like to add this to the Future Work section in this paper if possible.