On evaluating the literature.

Table of content

1. On evaluating literature

The aim of evaluating research is to offer a reasoned opinion as to the suitability, usefulness, relevance, and/or quality of other people's work, and of their work compared to your work. Such reasoned opinion is based on critically thinking about the content of the research, and using academic argumentation to justify your reasoning.

1.1. The nature of an evaluation

Evaluating scientific literature is about assessing the content and communication of science. The former is evaluation of scientific content, the scientific process, and scientific arguments and conclusions drawn from investigations. The latter is evaluation of the language (in a broad sense) used to communicate such investigations. In other words there are two things to consider when evaluating work:

1) *what it is we are evaluating*: hypotheses, theories, literature, methodologies, experimental procedures, results, interpretation of the results, arguments used to justify our conclusions, etc.

2) *how to evaluate*: we can evaluate work according to the i) advantages and disadvantages, ii) similarities and differences, iii) limitations, constraints, iv) strengths and weaknesses, v) gaps or omissions, and/or vi) significance, of the work involved. When writing about each of these aspects we need to write evaluation-type phrasing. This generally refers to making reasoned judgements, based on academic argumentation and appropriate language which identifies evaluation as being about advantages and disadvantages, similarities and differences, etc.

Note that not all of i) – vi) in 2) need be written about. It is up to the author, and yourself, to decide which of these will form part of your evaluation. So, when analysing a text for its evaluation simply use 1) and 2) as a reference to look back to in order to find out what aspect of the work the author is evaluating.

1.2. Evaluation criteria

Evaluating scientific literature can be done from two directions, where item 1) relates to evaluation of the scientific content: the validity of the scientific processes, arguments, and conclusions stated in a journal paper, and item 2) relates to evaluating presentation of the scientific work, specifically the language and argumentation used for providing reliable, high quality, truthful outcomes.

1) *directly*, by judging the science itself as described in the journal papers. This means that you need to be a specialist in the field. You need to
   a) understand current scientific theories which underpine the work described in the papers;
   b) understand how the research question or hypothesis of the particular scientific phenomenon being investigated relates to current theory;
   c) understading the literature used in the paper. Do they provide correct, accurate and reliable information.
   d) be able to judge whether or not the phenomenon being investigated is being investigated correctly: the nature of the apparatus, calibrations and testing; the way the experiments should be conducted (given the aims/theses) in order to judge the viability of the data collected (i.e. accurate versus inaccurate data) and the validity of the data collected (i.e. the data may be accurate but it doesn't address your intended hypothesis);.
   e) know how repeat or replicate the experiment yourself;
   f) to be able to suggests improvements/changes (if any) to the experimental methodology;
   g) know how to interpret their results;
   h) be able to decide whether or not their interpretations are correct, possibly validating or refuting methodologies, results and/or conclusions.
   i) Suggest remedies, solutions, areas of further work, etc.

2) *indirectly*, by judging the communication of the science, i.e. judging the struture of the paper, the quality of the argumentation, the coherence and logical development of the ideas in the journal papers, the cohesion of the language used. This means that you need to look at

   a) the abstract: this should be the most succint writing in the whole paper. So succint that it is possible to write only one sentence representing each section of the paper (except for the literature review section which does not need to be included in the abstract). Hence we should see the aim/hypothesis, methodology, results and conclusions (but no analysis or discussin or interpretation);

   b) the "purpose" of the paper, in other words whether or not the authors clearly state the aim/theses of their work. This is not only found in the abstract but also in the introduction. Is there a clear hypothesis, thesis statement, reason or justification for the work? Is it clear what it is that they are investigating? Is their aim/hypothesis testable and solvable (in other words, could you design an experiment in your mind which could test the thesis?);

c) the literature: Is there a sufficient description, summary and critique of the literature for the reader to see the context within which the authors' work lies? Is the link between the review, problem, and hypothesis evident? In other words, can you determine if the literature review is relevant to the work the authors are conducting?

d) the experimental methodology: is the apparatus suitably described? If necessary, are technical specification stated? Is there a good description of the experimental set-up (could you reproduce the experimental rig based on the authors' description)? Is the experimental procedure relevant to hypothesis, in other words will the experiment provide relavant data in support of hypothesis?;

e) the results (R), some useful questions being:
   - are the results viable/accurate. By this I mean is the experimental apparatus sensitive enough to produce results to 8 decimal places?
   - what is the quality of presentation of the results? For example, are tables properly constructed with only relevant data? Are graphs and diagrams clear and easy to "read"? Are they self-contained, in other words can they be read and understood without referring to the main body of the text (i.e. do they provide all necessary information as part of the visual)?

f) analysis and discussion (A.D.) as being (seemingly) plausible for the experiment being conducted and the hypothesis being tested. I say "plausible" because, unless you repeat the experiment yourself, you have to take it on trust that the data actually makes sense for the experiment being conducted (see the Blondot N-rays controversy of the early 1900s for a counter example to this).

   Some useful questions which can be asked are:
   - does the interpretation of the results make sense? Is the interpretation plausible? Is the argumentation plausible?
   - does the interpretation address the hypothesis?

g) conclusions: This should synthesise important finding. In other words, it should summarise the main contributions, and the relevance to the general discipline, under 1 single coherent commentary.

h) title, hypothesis, methodology, interpretation and conclusion: This whole-paper evaluation refers to the following:
   - can you follow the arguments put forward by the authors? Is the development of their ideas logical and coherent?
   - does the paper address the topic of the title?
   - is the experimental methodology appropriate for testing the hypothesis? Does the experimental methodology allow valid data to be collected (valid in the sense that the data is relevant to answering the hypothesis)? Is the experimental procedure described sufficiently well enough so that someone else could use it to repeat the experiment without having to guess or fill in any missing steps?
   - is the interpretation of the data plausible? Is it supported by the data? Does the evidence support the aim/thesis?
   - are the conclusions supported by the results and interpretation?

i) References: Are they properly listed? Are all the sources used in the literature review, and in the paper as a whole, listed in the references? Do other references seem appropriate?

j) the relevance or usefulness of other peoples' work in general: To evaluate any work as relevant one needs to know exactly the aims and objectives of one's own work so that one can compare this against those of other peoples' work.

   Some useful questions which can be asked are:
   - to what extent is this contribution significant or important or ground-breaking?
   - does the research fill in significant gaps in knowledge in the discipline?
   - does the research significantly advance the state of the art of the discipline?
   - does the research answer questions or solve a problem that has been studied for a long time but remained unanswered (for example, Fermat's last theorem)?

k) the relevance/usefulness of other peoples' work to your own work: How does one paper or a body of literature support you in conducting your own research?

Some pertinent question along this line are:

- is the information in the journal paper relevant to the work you are doing? The answer to this question is not black or white. Although a paper may be investigating a different hypothesis to your own, some information in the paper may still be helpful or useful to you.
- help you in your own investigation?
- support or refute your argument?
- provide information or ideas that you didn't know of before?
- is the information too basic or too advanced for what you need? Are you looking for something broader or more detailed? Irrelevant. Your write-up will be based on the fact that the work is pitched at the right level
- have you looked at a variety of sources before determining that these particular sources are the ones most relevant to your needs? If so, what is it about these particular sources that make them most relevant?

The above may look as if there is a lot to go through when conducting an evaluation. But with practice it won't seem that much. Some of the criteria above can be done fairly easily and quickly.

Since this course is not a science course, we will not be doing a direct evaluation of the scientific content of the literature. This would take us too deeply into the mathematics, physics, experimental methodology, and analysis of results relating to the scientific phenomenon we are investigating. We are therefore left with indirectly evaluating the scientific literature.

We can make it easier to conduct an evaluation by organising the above criteria, along with item 2) of section 1.1, into a table, as shown below.

	Abstract	Purpose /aims	Hypothesis	Literature	Experimental methodology	RAD	Whole paper
Similarities and differences							
Advantages and disadvantages							
Limitations or constraints							

Strengths and weaknesses							
Gaps or omissions							
Significance							
Relevance /usefulness in general							
Relevance /usefulness to own work							

Note that the table does not have to be filled. As previously mentioned it is up to the author to decide on the nature of his/her evaluation (i.e. which of the top row s/he addresses).

<u>Important comment</u>

We are not limited to just the criteria in the left column. At the time of wiring I found an extra criteria when analysing example 1 in section 1.5 which I have therefore included in the analysis of that example. So, if you find other criteria that qualify as evaluatory in nature then you can also use these. Make sure to make these new criteria explicit, or to explain why they qualify as evaluatory.

1.3. How I read papers

The way I tend to read papers is to scan through the paper looking at the title, reading the abstract, looking for the hypothesis again in the introduction and looking at the conclusions. This gives an overall picture of the paper, iis aims and outcomes.

If this paper interested me (or relevant to me) then I would read it in detail, specifically the detailed aspects of the author's aims, some/most of the literature review, the methodology, and then the analysis and discussion. I would read the results only in conjunction with the analysis and discussion since this latter usually makes reference to the results somewhere along the line.

However, with some philosophy papers or papers on the history of maths, I can't do any of the above since these papers are structured completely differently. They are basically written as extended essays. All I am left with is looking at the title and the abstract (if there is one, since some philosophy papers don't include abstracts), and then going straight into reading the paper.

1.4. The language of evaluation: Some simple examples

When writing an evaluation about scientific work the type of language we use has to address points 1) – 4) in the previous section, using language which addresses 1) – 2) of section 1.1. But at its most basic level a simple way to write evaluation-type language is to use adjectives, as shown in the left column of the table below. Compare this phrasing with that shown in the right column.

Evaluation-type language	**Non-evaluation-type language**
Providing excellent examples	Providing examples
Phlogiston is a defunct theory of combustion	Phlogiston is a theory of combustion
The discredited flat Earth theory posits that …	The flat Earth theory posits that …

Such adjectives stress the relevance, importance, quality, defect, issues, accuracy, etc., of the science according to one or more of the criteria 1) – 4) on the previous page. Note that the phrasing in the left column above does not need to be in adjective form. We could equally rewrite these evaluative phrases as follows:

- "Phlogiston is a theory of combustion which is now [or "has now been shown to be"] defunct",
- "The flat Earth theory posits that … This theory has now been discredited."

More generally, though, we need to be able to see phrasing as *intending* an evaluation. For example,

- "As a result of implementing the testing and calibration methodology we conclude that our data collection method is reliable." This is an evaluation comment about the reliability of the data based on the testing and calibration methodology. The implication here is that the methodology is of a sufficient quality (i.e. evaluating against a particular standard) to make the data collection reliable.
- "There is a crisis in science. It has become apparent that an alarming number of published results cannot be replicated." This is an evaluatory comment on the quality

of results obtained in the field of science in general (such comments tend to be seen a lot when it comes to evaluating the usefulness of significance testing in statistics, particularly in the field of medicine or social science).

- "Future work could include ...". This is evaluatory language because the authors have identified gaps or limitations in their own work. Such wording is standard in the conclusion part of academic papers.

All of the examples above are designed to illustrate a characteristic nature of evaluation-type language. The exact phrasings above may or may not be used directly in research papers. That is why it is important to develop experience in identifying evaluator-type language, along with specific terminology or wording, so that you can write evaluations easily.

1.5. Example evaluation language from the literature

Below are some examples of evaluation commentary from the literature. Note that not all the criteria of the left column of p5 need apply.

Example 1

The following text comes from Ivanov, Nikolov, Petrova (2014). The evaluation vocabulary or phrasing is highlighted underlined. Also, certain parts are annotated wherefrom you can refer to the relevant comments.

> "The hardest part of the experiment is measuring the diameter of the water jet at different heights under the tube[1]. This can be done photographically. We need to take a good close-up photograph of the water jet with the ruler alongside it[2]. From this we can actually measure the diameters. A proper scale can be established[3] by either using the measuring scale already in the picture alongside the water jet or by using d_0 which we have already measured very precisely. This, however, is not really needed[4] since we only need the proper value of d_0 to calculate v_0 – all other cross-section diameters are only needed in ratios to d_0 and these are the same whatever the unit of measurement (so long as it is the same for both diameters). We thus do not need to perform the cumbersome scaling calculations.[5]"

Notes

1.: Limitations/constraints: This sentence describes evaluating the degree of difficulty in measuring a certain property of the phenomenon being studied.

2.: Strength (of sorts): This is an evaluation comment because a good close-up photograph will be a strength when it comes to obtaining good quality data.

3.: Strength.

4.: Redundancy: This criterion isn't in the list on p5, but is still relevant here. Can you see why redundancy might apply here, and why it classifies as evaluator in nature?

5.: Weakness: Cumbersome work is always a weakness in research.

Example 2

The following text comes from Özmen (2004). The evaluation vocabulary or phrasing is highlighted underlined. Also, certain parts are annotated wherefrom you can refer to the relevant comments.

> "Students' misconceptions before or after formal instruction [on chemical bonding] have become a major concern[1] among researchers in science education because they influence[0] how students learn new scientific knowledge, play an essential role[2] in subsequent learning and become a hindrance[3] in acquiring the correct body of knowledge.
>
> In this paper some students' misconceptions on chemical bonding reported in the literature were investigated and presented. With this aim, a detailed literature review of chemical bonding was carried out and the collected data was presented from past to day historically. On the basis of the results some suggestions for teaching were made.[4]"

Notes

0.: Limitation (note that "they" refers to student misconceptions).

1.: Significance.

2.: Normally one might think this is a strength. But because this forms part of the sentence which continues onto 3. I would call this a weakness.

3.: Weakness.

4.: This whole paragraph addresses relevance to own work.

Example 3

The following text comes from Lee (2018). It is a highly technical paper with much density of language. Nevertheless, it is still possible to isolate the evaluator part of the text. The evaluation vocabulary or phrasing is highlighted underlined. Also, certain parts are annotated wherefrom you can refer to the relevant comments.

> "To date, a comprehensive description of plasma hysteresis and a physical understanding of the main mechanism underlying their bistability remain elusive[1], despite many[2] experimental observations of plasma bistability conducted under radio-frequency ICP excitation. This fundamental understanding of mode transitions and hysteresis is essential and highly important[3] in various applied fields owing to the widespread use of ICPs, such as semiconductor/display/solar-cell processing (etching, deposition, and ashing), wireless light lamp, nanostructure fabrication, nuclear-fusion operation, spacecraft propulsion, gas reformation, and the removal of hazardous gases and materials. If, in such applications, plasma undergoes a mode transition and hysteresis occurs in response to external perturbations, the process result will be strongly affected. Due to these reasons[4], this paper comprehensively reviews both the current knowledge in the context of the various applied fields and the global understanding of the bistability and hysteresis physics in the ICPs."

Notes

1.: Limitation or weakness.

2.: Strength or advantage (why?).

3.: Significance.

4.: This, and the sentence as a whole, addresses relevance to own work.

Example 4

The quotes below come from William Bechtel: "Scientific Evidence: Creating and Evaluating Experimental Instruments and Research Techniques", *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume One: Contributed Papers (1990), pp. 559-572.

- "The question of how scientific hypotheses and theories should be evaluated in light of evidence has been a central question in philosophy of science."
  Based on this statement one could say that hypotheses themselves are evaluated by subjecting them to experimentation. By the "quality" of a hypothesis we can say its

validity. Einstein theory of gravity as curved space(-time) was 'evaluated' by Eddington's solar expedition, calculating Mercury's orbit more precisely, and discovering black holes, etc.

- "At a given time many instruments used in a particular science are non-controversial. The techniques for using them are agreed upon so that, in a fairly routine way, scientists are able to generate evidence to settle empirical or theoretical disputes."
  This is an evaluation comment relating to the conventional types of scientific instruments used in experimentation. Most are "non-controversial" (as opposed to controversial). Evidence can be generated in a "fairly routine way" (as opposed to a problematic way).

- "There it is generally acknowledged that in order to test an hypothesis it is necessary to makes assumptions, generally referred to as auxiliary assumptions, about the test procedure itself. It is noted that if a prediction is not borne out, the auxiliary assumptions (treated now as additional hypotheses) may be at fault, not the hypothesis being tested."
  This is an evaluation on the relative merit of auxiliary hypotheses in order to see which are flawed, which can be saved, which have to be ditched, which new ones to introduce, etc., all the while keeping the main hypothesis intact because you have deemed (i.e. valued or judged or evaluated) this to be fundamentally sound.

- "Scientists generally do not appeal to theories about how their instruments work to develop or refine them, or to evaluate the reliability of the resulting evidence. Rather, they introduce and modify techniques by tinkering, that is, they physically alter the instrument or vary the way they perform the procedure. […] This raises the question of how scientists evaluate such tinkering: how do they determine which instruments and techniques are best and whether the resulting evidence should be counted as objective or as an artifact produced by the instrument or technique itself?"
  This comment is not an evaluation in itself. Rather it discusses evaluations made by others (scientists) with respect to the equipment they use.

- "The biologists who developed the techniques for using the electron microscope and the ultracentrifuge generally knew the basic theory about how these instruments worked. But they engaged in significant controversies over the procedures for using the instruments to study biological material and the reliability of the evidence so acquired. What is of interest is how these controversies were settled. I will suggest four factors:

(1) whether the technique yielded what appeared to be well-defined evidence (e.g., clear images in micrographs, sharp separation of fractions), (2) whether others were able to employ the technique with similar outcomes, (3) whether the evidence generated by the new procedure was consistent with evidence gathered using other techniques and (4) whether this evidence fit into plausible theories of the operation of the cell."

This comment is not an evaluation in itself. Rather, it address the way biologists went about evaluating the instruments they used and the evidence gathered therefrom: "controversies over the procedures for using the instruments" and the "reliability of the evidence" both imply the need for evaluation the instruments and the evidence. Items (1) – (4) then detail how to go about such evaluations.

Exercise

The text below comes from Palmroth et al (2018). Again, this is a highly technical paper with much density of language. Nevertheless, it is still possible to isolate the evaluator part of the text. As such, identify the evaluation vocabulary or phrasing, and identify which of the categories of the left column of the table on p5 these fall under.

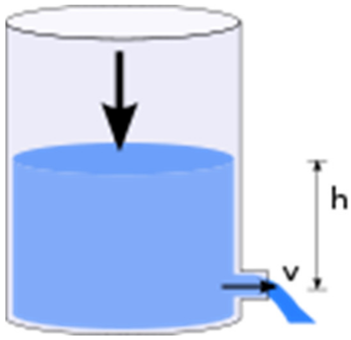
> "As standard verification tests for a hybrid-Vlasov system do not exist, the first verification effort of Vlasiator was presented in Kempf et al. (2013). A simulation of low-β plasma waves (where β is the ratio of thermal and magnetic pressures) in a one-dimensional case with various angles of propagation with respect to the magnetic field was used to generate dispersion curves and surfaces. These were then compared to analytical solutions from the linearised plasma wave equations given by the Waves in Homogeneous, Anisotropic Multicomponent Plasmas (WHAMP) code (Rönnmark 1982). Excellent agreement between the results obtained from the two approaches was found in the case of parallel, perpendicular and oblique propagation, the only noticeable difference taking place for high frequencies and wave numbers, likely as a result of too coarse a representation of the Hall term in the Vlasiator simulations at that time.
>
> In the work presented by von Alfthan et al. (2014), the study of the ion/ion right-hand resonant beam instability is another effort to verify the hybrid-Vlasov model implemented in Vlasiator against the analytic solution of the dispersion equation for that instability. The obtained instability growth rates were found to behave as predicted by theory in the cool beam regime, although with slightly lower values

which can be explained by the finite size of the simulation box used. This paper also discussed comparisons of results from the hybrid-Vlasov approach with those obtained with hybrid-PIC codes to underline that distribution functions are comparable albeit smoother and better-resolved with the former approach."

On a final note, it might be interesting to note that experiments act to evaluate the correctness of hypotheses.

1.6. Exercise: Writing an evaluation

In order to be able to show an evaluation of ones work compared to others one first has to have one's own work. Suppose, therefore, I am investigating the flow of water through a small hole at the bottom of a cylinder having thin walls.



My aim is to verify the phenomenon that the speed V of water from the hole is related to the height H of the water surface in the cylinder by the formula

$$V^2 = 2gH$$

where $g = 9.81m/s^2$. In order to verify this formula I decide to use a standard right-cylinder, as shown in the diagram above. A hole of 2mm is then drilled towards the bottom of the cylinder, this to act as the outlet. This outlet is initially sealed by plug, and water is poured into the cylinder. The plug is then released when we wish to start the experiment. When the experiments starts a special instrument is used to measure the speed of flow at the outlet for any given height. The resulting data will then be plotted, with the aim of confirming that V^2 is proportional to height H.

Having set the scene of my own research work you can now evaluate it against three other works, say of author A, author B, and author C who each use the different apparatus, but rigged up in the same way, and using the same experimental procedure.

- Author A uses thick walled cylinder, oil, and a 2mm hole at bottom of cylinder

- Author B uses thin walled cylinder, water, and a 4mm hole half way up cylinder.
- Author C uses thick walled cylinder, water, and 2 holes each of 2mm.

The exercises is for you to focus only on evaluating the experimental apparatus, and do so according to the some of the categories of the left column in the table shown on p5.

2. Evaluating the technicalities of the discipline

Here we go through two examples of evaluating the technicalities of a topic. The first example relates 5 different ways of presenting multiplication, and the second examples discusses means and medians in the context of statistics.

2.1. Example 1: On evaluating different ways to multiply two numbers

As research scientists you will ultimately end up evaluating experimental methods, data collection methods, results, and analyses/conclusions based on results. This is impossible to do unless one is a subject specialist. For example, how reliable is our data collection method? How do we know that one specific statistical test is the most appropriate one to use on our data for the hypothesis we are testing? What are the likely errors when conducting an experiment in this-or-that way? Hence the reason for us focusing on the broader aspects of evaluation illustrated in section 1 above.

However, we are all subject specialists in arithmetic. Because of this we can go about evaluating different approaches to multiplying two numbers some of which are shown below. Each has similarites and differences, advantages and disdavantages, limitations and benefits, etc. We can then think critically about the different approaches to multiplication, and write this up as an evaluation of such methods. This will be illustrated at the end of method 5.

Method 1: The usual way: Long multiplication

$$
\begin{array}{r}
325 \\
\times \quad 12 \\
\hline
650 \\
3250 \\
\hline
3900
\end{array}
$$

Method 2: Repeated addition

325 + 325 + 325 + 325 + 325 + 325 + 325 + 325 + 325 + 325 + 325 + 325 = 3900

Method 3: Split the multiplier into a binomial term and expand

Version a

$$
\begin{aligned}
325 \times 12 &= 325 \times (10 + 2) \\
&= 325 \times 10 + 325 \times 2 \\
&= 3250 + 650 \\
&= 3900
\end{aligned}
$$

Version b

$$
\begin{aligned}
325 \times 12 &= 325 \times (8 + 4) \\
&= 325 \times 8 + 325 \times 4 \\
&= 2600 + 1300 \\
&= 3900
\end{aligned}
$$

Method 4: Halving and doubling.

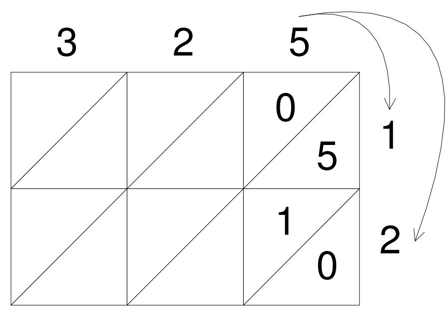
For 14 × 12, halve one number (ignoring any remainder) until you get to 1, and double the other number. Then add all doubled numbers that lie across odd halved numbers:

14	12
7	24
3	48
1	96

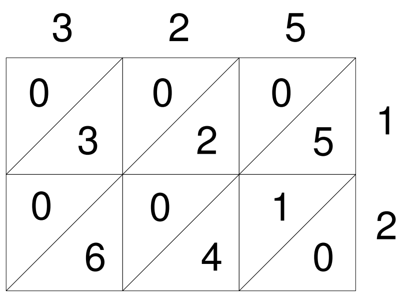
Since all numbers in the halved (left) column are odd we add up all the numbers in the right column. So 14 × 12 = 24 + 48 + 96 = 168.

Method 5: Grid multiplication

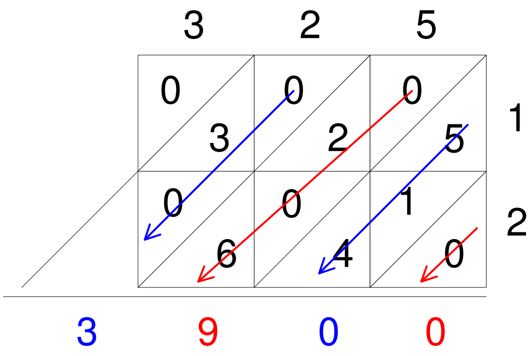
Place one number to be mutiplied on top of a grid and place the other number to the side of the grid as shown below. Then multiply each digit with each other digit and place in the relevant triangle. For example



So we get



Finally, add along the diagonals, carrying as needed:



Now for the evaluation

Here we will go through an evaluation based on the similarites and differences, advantages and disdavantages, limitations or constraints and significance of these methods, …? So,

1) Method 1 and method 5 are similar because because the lay out both numbers in hundreds, tens and units. However, method 1 differs from method 5 because this latter method lays out one number horizontally and another number vertically, whereas method 1 lays out bot numbers horizontally. The advantage (if it is an advantage) of method 5 is that the child doesn't have to indent subsequent rows of multiplied digits as is the case in method 1. The required indentation is achieved in method 5 by the diagonals of the rectangle, this being given to the child beforehand. The disdavantage of method 5 is that the layout works only for multiplying two numbers not three of more numbers. How would one layout the following in one grid form?

```
     325
×     12
×    713
```

In that case the grid form would need to be performed twice, once on 325×12, then on 3900×713.

2) There are no similarities between method 2 and method 4. Method 4 is completely different from method 2 since it relies only on multiplying and dividing by 2, and taking account of the remainder of such a division. Method 2, on the other hand, has no multiplication in it and does not deal with remainders (which is an aspect arising from multiplication). Both these methods has a severe disadvantage/limitation because both repeated addition and repeated halving/doubling take far too long when dealing with large numbers. The only advantage I can see is the ultra simplicity of both methods. This certainly might be useful from a pedagogicalperspective when teaching a child how to multiply for the first time, and helping him/her see the arithmeitc connections between the two methods and normal multiplication.

Note that there are no gaps or omissions to the methods since all steps of each method is clearly shown and necessary in order to arrive at the answer. Also, from a mathematical point of view I would say that none of these methods are significant. Once the process of multiplication has been fully defined (usually method 1) then all other methods bring nothing new to the mathematical nature of multiplication. However, from a pedagogical point of view, any of methods 2) to 5) could be significant. A child having difficulty learning or understanding method 1 might benefit greatly from one of the other methods and gain an understanding of the nature of multiplication.

Comparison between other combinations of methods could also be done. This is left as an exercise.

Exercise: Below is another approach to multiplication. By comparing and contrasting this method with those above, write an evaluation of this method.

$$97 \times 91 = \mathbf{8827}$$

100 – 97 ↓ 3   +   100 – 91 ↓ 9   =   12 → 100 – 12 = **88**

$$3 + 9 = 12$$

and

$$3 \times 9 = \mathbf{27}$$

For your own interest, the following (which does not need to form part of your evaluation but can do so if you wish) justifies the mathematical validity of the above process:

$$\begin{aligned}
97 \times 91 &= (100-3)(100-9) \\
&= 100^2 - 100 \times 9 - 100 \times 3 + 3 \times 9 \\
&= 100^2 - 100(3+9) + 3 \times 9 \\
&= 100[100-(3+9)] + 3 \times 9 \\
&= 100 \times 88 + 27 \\
&= 8827
\end{aligned}$$

2.2. Example 2: On evaluating approaches used to find averages of data in statistics

Here we will compare and contrast the use of means and medians as ways of finding the average of data, as well as standard deviations and interquartile ranges as ways of finding the spread of data. As such, consider the following text (which I invented myself) as a primary piece of writing:

> When it comes to a basic analysis of data Smith (1980) uses the mean as his measure of central tendency. He explains that the advantage of this is that, because all his data values are used in finding the mean, taking the mean for different samples of a population tends to give similar results. This indicates that the mean is robust, i.e. it resists very well any fluctuations between different samples.
>
> On the other hand, Jones (1990) makes use of the median as his measure of central tendency. This is because the median does not rely on arithmetic but on ordering the data from smallest to largest and then identifying the middle value. This middle value is a form of average because it represents the most common value among the data.

Evaluation

> Both Smith (1980) and Jones (1990) focus on measures or central tendencies. However, the fail to discuss measures of spread such as standard deviations and inter-quartile ranges. Measures of spread are extremely important, and crucial when it comes to performing statistical inference on continuous data. As a consequence, the absence of these latter measures from their discussion is a serious flaw.

Despite this, there are pros and cons to using either means and medians. The problem with using the mean is that it is sensitive to outliers. The further the outliers(s) the more it will affect the mean, resulting in a value of the mean which is not representative of the "middle" of majority of the data. So one single value can significantly skew the value of the mean away from the most representative average.

Now, it is true that using the median has the advantage of overcoming the problem caused by outliers when using the mean, but the disadvantage of its use is precisely that it is not calculated arithmetically (and therefore does not use each data value). This results in the median being easily affected by the type of sample we take from the population, and is such a significant drawback that it is not used in statistical inferencing.

Exercise: Is the text above an evaluation? If not, why not? If so, why? Identify the language used in the text which makes it an evaluation.

3. References

Ivanov, D., Nikolov, S., Petrova, H. (2014), "Testing Bernoulli's law", Physics education 49 (4), July 2014.

Lee, H-C (2018): "Review of inductively coupled plasmas: Nano-applications and bistable hysteresis physics", Appl. Phys. Rev. 5, 011108 (2018).

Özmen , H., (2004): "Some Student Misconceptions in Chemistry: A Literature Review of Chemical Bonding", *Journal of Science Education and Technology*, Jun., 2004, Vol. 13, No. 2 (Jun., 2004), pp. 147-159

Palmroth, M. et al (2018): "Vlasov methods in space physics and astrophysics", *Living Reviews in Computational Astrophysics* (2018), vol 4, issue 1.