

Thinking about Scoring

On this page I outline two proposals:

- Application Learning Efficiency (*ApplEF*) and *Application Effectiveness* (*ApplE*) - An approach that produces a single score for applications which combines the per step approvals and the overall facilitator assessment but is a bit awkward because it combines percents and what can be viewed as a letter grade.
- Facilitator Assessment and Comprehension Score Concept - An approach based on the idea that we can get **some** comprehension information from Applications also.

Application Learning Efficiency (*ApplEF*) and *Application Effectiveness* (*ApplE*) Proposal

Lets abbreviate "Application Learning Efficiency" as "*ApplE*", and "*Application Effectiveness*" as "*ApplEF*".

- *Appl_EF* is the facilitator response encoded as 0=strongly disagree 1=disagree 2=agree 3=strongly agree.
- *Appl_LE* is the discounted correctness $\text{avg}(\text{discount}(\text{attempts}))$ over all questions in the topic.

Exactly what *Appl_EF* means is entirely determined by the text of the statement the facilitator is asked his opinion of. That I think has already been settled:

This application provides evidence that this learner can effectively apply the concepts in their actual work.

Exactly what *Appl_LE* means is determined by the choice of discount function. If a learner takes 10 submissions to get a question right, I think that should be considered unreasonable. In fact 5 is probably excessive, while 3 is probably reasonable. Maybe its acceptable to just take off 10% for each attempt.

1 attempt=> 100%, 2 attempts=> 90%, 3 attempts=> 80%, 4 attempts=> 70%, 5 attempts=> 50%
6 attempt=> 40%, 7 attempts=> 30% 8 attempt=> 20%, 9 attempts=> 10%, 10 or more attempts=>0%

But maybe you want something non-linear, for example.

1 attempt=> 100%, 2 attempts=>95%, 3 attempts=>85%, 4 attempts=>70% 5 attempts=>50%
6 attempt=> 20%, 7 or more attempts 0%.

And maybe you want to be more severe:

1 attempt=> 100%, 2 attempts=>90%, 3 attempts=>70%, 4 attempts=>40% 5 or more attempts=>0%

One you decide on the above the bigger question is: what do you want to DO with these 2 scores ?

The current proposal is to combine them into a single "application score".

The difficulty with that is *AppLEF* is a "letter grade" while *AppLE* is a percentage. We need to set up a table of equivalences:

letter grade:	0 (weak)	1 (moderate)	2 (good)	3 (expert)
<i>AppLEF</i> :	"strongly disagree",	"disagree"	"agree"	"strongly agree"
<i>AppLE</i> :	0-70%	70-80%	80-90%	90-100%

Then we combine the letter grades just by averaging their numerical values, and rounding back to an integer. This will give us an application score which is expressed as a letter grade.

If you instead want an application score which is expressed as a percentage, maybe you can use *AppLEF* to modify *AppLE* according to rules something like:

if <i>AppLEF</i> ="strongly agree"	combined score is the average of <i>AppLE</i> and 100%
if <i>AppLEF</i> ="agree"	combined score is the same as <i>AppLE</i>
if <i>AppLEF</i> ="disagree"	combined score is 80% of <i>AppLE</i>
if <i>AppLEF</i> ="strongly disagree"	combined score is 50% of <i>AppLE</i>

So, let's take a sample problem: A learner, Jen, has gone through the two applications in Album A, Application A1 and A2. She got all 5 steps of A1 right on the first go round. With A1, she got the first two steps of the 10 steps in A2 wrong the first time and the rest correct and on the second iteration she got all but the first step right and on the third she had them all right. Let's also say that the facilitator's Likert response on A1 was "disagree", but that on the third attempt, his response on A2 was "strongly agree".

I will use the last discount function above in this example.

So, one would score for her Application A1 as follows:

Since she got all 5 steps approved on the first attempt she has *AppLE*= 5/5= 100%, but she only got *AppLEF*=1. If we go with letter grades, she has $\text{avg}(3,1)=2$, a grade of "good" over all. If we go with percentages, her 100% gets reduced to 80% following the above rules.

And A2 as follows:

- Here she has all but 2 steps right on the first attempt +8.00
- She got one of the remaining on the second attempt +0.90
- and one on the third attempt +0.70, so $\text{Appl_LE} = 9.6/10 = 96\%$
- Moreover she has $\text{Appl_EF}=3$ ie "expert".
- for a letter grade, appleLE is in also "expert" range so overall she scores "expert"
- for a percentage we average that 96% with 100% producing 98%.

One might then agg Jen into her class as follows:

- First we have to give her a score on the whole album, combining Application A1 and A2.
- Since Appl_EF is only provided per-topic, we simply average her scores on each topic (assuming they are weighted equally).
- For a letter grade "good" + "expert" $/2 = 2+3/2 = 2.5$ (do you round up or down ?)
- For a percentage $(80\% + 98\%)/2 = 89\%$

To "aggregate her into her class" we just include her score of 89% in the average along with her class mates. We can also provide a class average on each topic independently.

Facilitator Assessment and Comprehension Score Concept

Application Effectiveness (the facilitator Likert response) should stand on its own as "application score" while Application Learning Efficiency should either be a component of "comprehension score" or also stand as an independent score.

Then you get a comprehension score which is a percentage and an application score which is a "letter grade" and no arbitrary conversion are necessary.

Moreover "Application Learning Efficiency" is a straightforward generalization of comprehension score to application topics.

In each case the learner gets a score (bet 0-100%) on each question and these are averaged to produce a score on the whole topic.

A comprehension topic may only be submitted once, and the score on each question is EITHER 100% or 0% (correct or not).

An application topic must be submitted as many times as necessary for all **steps** to be considered correct, but the score on each question decreases with the number of submissions.

In each case it is about the learner's ability to read and digest the learning materials and the questions and come back with an appropriate response.

The Application Effectiveness is different. It a subjective human opinion as to how well the learner would be able to apply the concepts in the real world as opposed to a test-taking environment. It is worth considering if it is more useful to combine the scores that are more structurally similar and keep Application Effectiveness separate.

```
completion score= (learning_topics_viewed + comprehension_topics_submitted + application_topics_approved)/ all_topics_published
comprehension score for comprehension topics= sum(correct_each_question)/questions_published
comprehension score for application topics= sum(discount(attempts_each_step))/application_steps_published
```

By analogy with "completion score" which includes all 3 type of topics, we might define "comprehension score" to include both comprehension topics and application topics.

This could be done just by averaging the two comprehension scores above, which would weight topics equally, or by summing the numerators and denominators separately and then dividing, which would weight the questions and steps equally.

We can then add a new score for the Likert response, which unlike the others is a letter grade. I tend to like the name "facilitator assessment" or "assessment score" for this.

```
facilitator assessment= round(sum(facilitator_response_each_application)/application_topics_published).
```

To rework the example of Jen above, would score for her A1 Application as follows:

comprehension= applLE= 100% application= ApplEF= 1 (moderate).

And A2 as follows:

comprehension= applLE= 96% application= ApplEF = 3 (expert)

One might then aggregate Jen into her class as follows:

- First we have to give her a score on the whole album, combining Application A1 and A2.
- comprehension is aggregated per-question (ratio of avgs) $(5+9.6)/(5+10) = 14.6/15 = 99.3\%$
- application is aggregated per-topic $(1 \text{ [moderate]} + 3 \text{ [expert]})/2 = 2 \text{ [good]}$

To "agg her into her class" would just average Jen's 99.3 in with the comprehension scores of her classmates and average Jen's 2=[good] in with the application scores of her classmates and round to an integer.