Assessing Proofs With Rubrics: the RVF Method

David E. Brown Shayla Michel

Utah State University
Logan UT, 84322

March 15, 2010

Abstract

We present an easy-to-implement 3-axis rubric for the formative and summative assessment of open-ended solutions and proofs. The rubric was constructed for the use on the written work of students in a Discrete Mathematics class at a research-oriented university, with the following in mind: (1) To aid in the efficiency and consistency of assessment of proofs and open-ended solutions, with the possibility of being comfortably implemented by an undergraduate assistant; (2) To provide the simultaneous formative and summative assessment of the students' written work. Thus, the questions we address are: How can we foster good technical writing skills in a way that improvement can be measured? How can large amounts of written work be processed and assessed so that summative and formative judgments are passed but without much time used by the instructor/professor/TA? The axes we use are labeled *validity*, *readability*, and *fluency*, corresponding to (respectively) correctness of calculations and deductions, the ease with which the solution or proof can be read, and the extent to which a student is able to use and communicate via the technical notions relevant to the problem or proof — for example appropriateness and correctness of notation. The rubric format is communicated to the students and discussed in class before any written work is assessed. The rubric has been implemented by professors and teaching assistants only after being trained in its use.

1 Introduction

In an attempt to increase the efficiency of summative assessment and the efficacy (and efficiency) of the formative assessment of written work of students in a third-year university Discrete Mathematics class whose topics include Set Theory, Logic, Enumerative Combinatorics, and Graph Theory the authors addressed the following questions while examining past student work and contemporary Mathematics textbooks.

- What constitutes good proof writing, and can the characteristics be identified to the extent that a metric be created to rank proofs?

- Characteristics mentioned above found or not, how can good proof writing in particular, and good technical writing in general, be fostered in an efficient way?

We note that requisite courses for our Discrete Mathematics course do not entail formulating or writing proofs, hence we have the responsibility of training the students to do so. Also note, while issues of students' notions of what constitutes a proof are clearly relevant, as treated in [2], we treated them with generous disregard as far as this experiment is concerned. It may however be relevant to note that the first author (henceforth Dr. B) presents, discusses and dissects fallacious proofs and correct proofs and also occasionally suggests template-style methods as in [4] in conjunction with the course content. In addition Dr. B engages students in a discussion about the variable meaning of 'proof' in the contexts medicine, law, science, Mathematics, and "everyday life." Students are asked to associate the terms deduction, inference, evidence, and precedent with their notion of proof in X, where X is one of the contexts listed. The results of these discussions has been observed and recorded by Dr. B but is beyond the scope of this paper.

We now mention the expectations of the students in our Discrete Math class and how they are communicated.

2 Dr. B's Expectations

The expectations of the students in this Discrete Math class are communicated in part by a discussion during the first or second class meeting revolving around examples of what Dr. B categorizes as "good" and "not-so-good" responses to prompts. The following prompt is presented by Dr. B and it is noted that there is not an expectation that any progress be made on the problem posed or even a complete understanding of the problem, but it is the readers' (students') reactions to the example responses taken from students that are the focal points.

Prompt: Let $a_1, a_2, \ldots, a_n$ be real numbers. Determine a recurrence relation for the number of ways to place parentheses in order to carry out the computation $a_1 \times a_2 \times a_3 \times \cdots \times a_n$ without permuting any of the numbers.

Here are two examples of responses from students which are presented and discussed.

Example 1. All punctuation, grammar, symbols, and spacing are represented as I found them.

> Parentheses problem:
> Note: Calculators can only multiply one number at a time.
>
> $a_n =$ to the number of ways to multiply $(x_1)(x_2)(x_3)\ldots(x_n)$.
> $i =$ the first choice to multiply. $1 \leq i \leq n-1$ ex: $((x \cdot x \cdot)x \cdot x \cdot x)$
>
> i can be anywhere, not necessarily the item on the left

So, $a_n = a_1 a_{n-1} + a_2 a_{n-2} + \cdots + a_1 a_{n-1}$

$a_1 a_{n-1}$ needs to be multiply because after you multiply a_1, you must do the rest of the multiply. So after a_i, a_{n-1} must be included. Therefore $a_n = a_i a_{n-i}$.

Example 2. Again, what I present is exactly as the student presented it.

Find a recurrence relation for the number of ways to place parentheses (i.e., to associate numbers) to compute $x_1 \times x_2 \times x_3 \times \cdots \times x_n$.

Claim: Let a_n be the number of ways to carry out the multiplication $x_1 \times x_2 \times \cdots \times x_n$. The number of ways to place parentheses in order to multiply $n \geq 1$ numbers can be determined by the recurrence relation $a_n = \sum_{i=1}^{n-1} a_i a_{n-i}$, $n \geq 2$, with initial condition $a_1 = 1$.

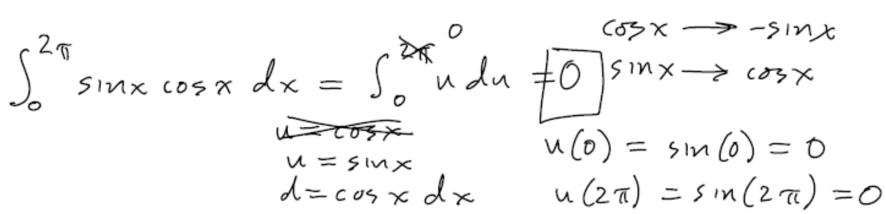
Proof: We can determine a_n by first splitting the n numbers into two groups, which will in turn be recursively divided into smaller groups to be multiplied together. If the first group contains a single number, we must multiply it with the product of the other $n-1$ numbers, which can be done in $a_1 a_{n-1}$ ways. If the first group consists of two numbers, the product of these, which can be computed in a_2 ways, must be multiplied by the product of the other $n-2$ numbers, which can be done in a_{n-2} ways. In general, if the first group consists of i numbers, then their product must be multiplied by the product of the other $n-i$ numbers; this can be done in $a_i a_{n-i}$ ways. Since i ranges from 1 to $n-1$, the total number of possible ways to complete the multiplication is $a_n = \sum_{i=1}^{n-1} a_i a_{n-i}$, and clearly we can and should take $a_1 = 1$ as our initial condition. ■

Dr. B asks students to make any observations about the two responses, to compare them, to make comments about what is or is not notable, or even to praise or criticize them. In each of the past 3 semesters this discussion has been conducted, there is always a student who remarks something like "I think I understand the problem better after reading the second response."

Dr. B emphasizes that it is responses of the second type that are sought, and then another more familiar prompt is considered:

Prompt: Please evaluate the definite integral $\int_0^{2\pi} \sin x \cos x \, dx$.

Example 3. Here is what a student presented, rendered as closely as possible to the original.



At this point, before presenting the next example Dr. B asks the class to comment on this problem and typically, since perhaps the students are now conditioned to find flaws in the first example given, they criticize the above example. Dr. B then asks “OK, if you were grading this, and it were worth 10 points, how many would you give it?” Almost always the informal poll indicates that nearly everyone would give the response full credit. Then Dr. B presents the following response to the same prompt.

Example 4. Another example of a student response.

In order to evaluate the definite integral $\int_0^{2\pi} \sin x \cos x \, dx$, I'll perform a variable substitution (as Dr. Brown wants us to call it), or a so-called “u-substitution” as our text book calls it, using $u = \sin x$, and, therefore, $du = \cos x \, dx$. To complete the substitution, we change our limits of integration with respect to the substitution: $u(0) = 0$ and $u(2\pi) = 0$. So, we have $\int_0^{2\pi} \sin x \cos x \, dx = \int_0^0 u \, du$, and this obviously evaluates to zero by one of the theorems in our text which says $\int_a^a f(x) \, dx = 0$ for any function f and any $a \in \mathbb{R}$.

At this point Dr. B asks for comparisons of the two responses and asks students to determine which student has indicated more clearly an understanding of the relevant content behind the solution to the problem in the prompt. Typical responses, although not all, indicate that the second response indicates a better understanding. Dr. B, unless students do this, argues that the second response indicates a clear comfort with the material (clear enough to poke fun at Dr. B via “... a variable substitution (as Dr. Brown wants us to call it), or a so-called ...”) and evidence that they have read the text and understand it well enough to cite apply its theorems in the appropriate places. Dr. B always poses the rhetorical question “If you gave these two responses the same score, would yo feel as though you have accurately indicated achievement through this score?” Dr. B argues that were it not for some enculteration that occurs in Mathematics classes, the first response for the integral prompt is not readable: You cannot follow the rule top-to-bottom and left-to-right in order to read the response. Dr. B claims “There is some accepted lack of clarity and benefit of the doubt bestowed upon students which allows a poor response as this to be scored highly.” This (possibly) hyperbole is among the ways the expectations are communicated.

3 Assessment: Pre RVF

The account above of one of the mechanisms by which expectations are communicated is now incomplete – expectations are now complemented with an explanation of the three categories of the RVF rubric, which will be given below. in this section we summarize how students' responses were assessed before the advent of the RVF rubric.

The first prompt associated to Example 1 and Example 2 above was accompanied by an objective-based rubric as follows. That is, each written assignment would be returned with a sheet of corresponding rubrics attached (an example is given in Appendix A). Here is the one that was used for the placing parentheses prompt:

Exercise 4. Placing parentheses. Let p_n be the number of ways to place parentheses.

- Author clearly displays a recurrence relation with initial conditions 0 1 2
- Author gives a cogent argument for the recurrence 0 1 2
- Author's recurrence is correct; it is $p_n = \sum_{i=1}^{n-1} = p_i p_{n-i}$, for $n \geq 2$ 0 1 2
- The initial conditions given are correct: $p_0 = 0$ and $p_1 = 1$. 0 1

The score 0 is given to an objective if it is clearly not met, 2 if it is clearly met and 1 if it is arguable but not clear. If the objective is arguably Boolean, then 0 or 1 is the score depending on whether the objective was *clearly* met or not. This scheme saved Dr. B time and was a tool that could be shared among teaching assistants and graders, but when a response fell outside of the anticipated one, it was inefficient and unfair. Other deficiencies of this rubric, as far as the authors are concerned, are that it is not versatile, and by its design cannot be shared with students in advance so that they better-know what is expected of them. Indeed, when queried, many students said something to the effect that if their scores were not all 2s then they would *not* use the feedback for improvement (i.e., not use it as a formative assessment), but merely conclude something along the lines of "I didn't do what Dr. B wanted" and their subsequent efforts would be independent of the feedback.

4 Revision of Assessment Process: The RVF Method

The following 3-axis rubric was developed which, at the time of writing this paper and based on recorded scores and on queries to students, is working in a way closer to the intentions of the authors. The rubric for the above prompt, or any prompt, under this rubric scheme would be

- Author's solution is readable 0 1 2 3
- Author's arguments and calculations are valid 0 1 2 3
- Author evinces fluency 0 1 2

The ranges of scores on the first two categories are augmented to match the weight given to the prompt via the first rubric, but is of course unnecessary and is only presented to show that flexibility of weight per problem is possible. However, we believe that a scale that is close to Boolean is best for consistency's sake and that 3 is an effective and efficient maximum for each category *readability*, *validity*, and *fluency* which we now informally define through what is presented to students before any written work is assessed. We note beforehand that *fluency* has been difficult to define, and is probably best "defined" by examples, but what is listed below is what is now distributed with the syllabus and is discussed on the first or second class meeting in conjunction with that given in Section 2.

> Your written responses will be assessed via attention to the following categories, which will be explained below: *validity*, *readability*, and *fluency*. Assume that an equal weight will be applied to each category that is applicable, noting that readability and fluency are always applicable whereas validity may not be.

Readability. Your written work will be examined for readability. Obviously, if your written work is not readable it cannot be assessed, but since the ability to communicate Mathematics is a focal point for this class, special attention will be paid to this quality. Strings of equations with no explanation or motivation do not constitute a readable response, and although a carefully chosen, motivated, and explained figure may be worth 1000 words, 1000 figures in and of themselves will be worth nothing.

Fluency. Mathematics is a concise and precise language and I wish to enhance your fluency with this language. Therefore, part of every assessment will focus on your ability to incorporate correct, established notation and terminology into your written work. An example of a lack of fluency is the use of "top" and "bottom" for the numerator and denominator of a fraction. Another example which is evidence of fluency is using "Let f be a real valued function of a real variable" interchangeably with "Let $f : \mathbb{R} \to \mathbb{R}$". Other things which are definable perhaps only by an exhaustive list of examples will be assessed as well.

Validity. If a solution requires calculations or entails a string of deductions, the extent to which your solution is valid will be analyzed. Validity corresponds to the discernable extent to which your method used is appropriate, your calculations are correct, and your deductions follow the rules of logic.

Several factors motivated the creation of these three categories, one of which was parsimony, but others we'll now explain. We created these categories after careful examination and discussion of dozens of student papers left behind or not collected after final exams. We identified written work that we thought of as "good" and that which we thought of as "bad." Then we attempted to define what exactly was good and bad about the work. Identifying the "bad" work was straightforward: errors in logic, and misuse of notation which often rendered the work unreadable (whence the readability category). It was more difficult to define what good about the papers and in particular, it was difficult to pinpoint what was *really good* about some of the paper we though exhibited really good, exceptional, work. This is where the fluency category came from. In some cases, we thought of it as elegance, or even something in the written responses which indicated "Mathematical maturity." Whatever it was, no student paper which we thought of as exceptionally good had inappropriate notation or notation used clumsily with standard English. We found that many, maybe even most, papers arguable exhibited validity, especially when given the *benefit of the doubt* as discussed in relation to the integral prompt above. In addition the second author (henceforth Miss M) gave valuable insight on how students interpret the rubrics when they are returned with their papers. It was posited that the simplicity of the categories coupled with a continued discussion of them as the class progressed would serve students well.

To further justify the three axes and in addition illustrate their independence, we identified student works which we believe warrant scores, under the format (Readability, Validity, Fluency), as (high, low, low), (low, high, low), and (low, low, high). In order to conform to the space constraints of this paper, and suppress the need to give a course's worth of definitions and notation, we give general qualities of papers we categorized into the three just mentioned. In addition, we focussed our attention on an assignment on tournaments. The necessary background on tournaments and examples from that context are given after the general descriptions.

Low Fluency, High Validity and Readability. These papers relied heavily on metaphor or a picture they would draw for their explanations and arguments. There seemed to be a reluctance to use the accepted definitions and terminology even when they would have clearly provided an economy of thought and expression.

Low Validity, High Fluency and Readability. These examples were less abundant than the others. Most of the papers containing responses which we placed in this category used arguments couched in examples, but with clearly explained figures, appropriate and smoothly used notation, and a narrative describing the examples which made them very readable. We believed that these types of papers were associated to students who typically perform well, but simply could not completely solve that particular problem. We did not have enough data to validate this hypothesis. Other examples were vacuous in the sense that their response was clearly simply an attempt at saying *something*, but strictly speaking they were readable and we had no way to argue that fluency was absent.

Low Readability, High Validity and Fluency. These examples were abundant when the content behind the prompts involved calculations, as in our unit on set theory and in particular asking questions about sets such as $\{(a, b, c) \in \mathbb{Z}^3 : a^2 + b^2 = c^2\}$ and $\{(p, q, r) \in \mathbb{Z}^3 : p = x^2 - y^2, q = 2xy, \text{ and } r = x^2 + y^2, \text{ where } x, y \in \mathbb{Z}\}$. These types of responses were mostly strings of equations with little or no narrative tying them together – "hieroglyphics" as we came to call them. To be a little more clear, the first response to the integral prompt from Section 2 would have been in this category were "$\frac{d}{dx}\cos(x) = -\sin(x)$" and "$\frac{d}{dx}\sin(x) = \cos(x)$" displayed for $\cos(x) \to -\sin(x)$ and $\sin(x) \to \cos(x)$, respectively.

4.1 Low Fluency Examples from Problems Regarding Tournaments

An *n-tournament* is a directed complete graph on n vertices; every edge of a complete graph is given a direction, equivalently the unordered pairs of vertices which constitute the edge set in a graph become ordered pairs. Indicating that there is an oriented edge *from* vertex u *to* vertex v is done by writing $u \to v$. A *queen* in a tournament is a vertex q with the property that for every other vertex z, $q \to z$ or there is a vertex x for which $q \to x$ and $x \to z$. Note that if the relationship $u \to v$ is not present in a tournament and u and v are actually distinct vertices, then $v \to u$ is necessarily in the tournament. With T an n-tournament, and v a vertex from T, $d^+(v) = |\{u : v \to u\}|$, and $d^-(v) = |\{u : u \to v\}|$. Here are specific examples from the mini-unit on tournaments:

Prompt: *Prove whether the following statement is true or false. In an n-tournament, it is possible to have both a vertex x with $d^+(x) = n - 1$ and a vertex z with $d^-(z) = 0$.*

Response: If vertex x has $d^+(x) = n - 1$, then it has an arc from it to every other vertex except one. That node is itself. Since x is sending out to every node, z must have an in-going arc and cannot be $d^-(z) = 0$

Prompt: *Please prove that every n-tournament has a queen.*

Response: Assume n is the tournament with no queen. Let H be defined as the vertex with the highest number of out-degrees. By definition of a queen it is known that a queen beats everything directly or indirectly. Based on our assumption, H cannot be a queen and thus at least one other vertex must beat H. Let vertex B be defined as that vertex that beats H. Let P be defined as all of the out-degrees of H. Accordingly, the number of out-degrees of B must include those same P arcs, or else H would beat B indirectly and thus be a queen. But this is not a queen tournament, so B's out-degree is P + W. And the number of out-degrees of B is one more than H. H is defined as the vertex with the most out-degrees but F has one more out-degree than H, making it the vertex with the most out-degrees. This leads to a contradiction because both F and H can't both have the highest out-degree. The assumption of having a tournament with no queen leads to a contradiction. Thus we have proved that every tournament has a queen and additionally vindicated what they say about making assumptions.

5 Conclusions

Dr. B records histograms of all scores on all assignments and within the first three assignments there was a consistent shift of the largest of the bin sizes to the high end of the scale. Other interesting observations are that more and more students are turning in typed and very well-presented assignments and seem to be taking more pride in their work. These are of course merely observations and Dr. B has begun to interview students in order to determine what they actually do with the information from the rubrics. As for the goal of efficiency, Dr. B spends about 1/2 of the time assessing students work than he once did and is even comfortable with sharing the burden of the assessment with undergraduate teaching assistants, but only after they have been trained in the use of the rubric and the meanings of the categories.

Miss M has suggested a combination of the objective-based rubrics and the RVF rubric for certain assignments. Miss M believes the objective-based rubrics can communicate exactly what was done wrong as far as the technicalities are concerned.

References

[1] Selden, J. and A. Selden, Teaching Proving by Coordinating Aspects of Proofs with Students Abilities, Tennessee Tech. U. Technical Report No. 2007-2, October 2007.

[2] Selden, J. and A. Selden, Overcoming Students' Difficulties in Learning to Understand and Construct Proofs, Making the Connection: Research and Teaching in Undergraduate Mathematics, Ed: M. Carlson and C. Rasmussen, MAA Notes no. 73 (2008), 95 – 110.

[3] Herbst, P. and C. Brach, Proving and Doing Proofs in High School Geometry Classes: What Is It That Is Going On for Students? *Cognition and Instruction*, 24 (1), 73122, 2006, Lawrence Erlbaum Associates, Inc.

[4] Smith, D. , M. Eggen, and R. St. Andre, *A Transition to Advanced Mathematics*, 5 ed. (2001) Thomson Brooks/Cole.

Appendix A: Rubric from Pre-RVF

MATH 3310
Grading Rubric for Homework Assignment #8: Poor Proof Pinpointing, and 2,3,4,5,6 from *Notes on Recurrence Relations.*

Author:

Exercise 2. Penny-pinching puma: (a) \$1000 for n years, (b) \$100 at the end of each year for n years. Let m_n be the amount of money after n years.

- Author clearly displays the recurrence and initial conditions[1] for (a) and for (b) — 0 1 2
- The initial condition $m_0 = 1000$ is given for (a) and $m_0 = 0$ for (b) — 0 1 2
- Author displays $m_n = 1.08m_{n-1}$ for (a) and $m_n = 1.08m_{n-1} + 100$ for (b) — 0 1 2

Exercise 3. n balls into k boxes with 2 to 4 balls in each box. Let $\left[{n \atop k}\right]$ denote the number of ways to distribute n balls into k boxes with the desired constraints.

- Author clearly displays a recurrence relation and initial conditions — 0 1 2
- Author gives a cogent argument for the recurrence relation — 0 1 2
- The recurrence displayed is $\left[{n \atop k}\right] = \left[{n-2 \atop k-1}\right] + \left[{n-3 \atop k-1}\right] + \left[{n-4 \atop k-1}\right]$ — 0 1 2
- Author's initial conditions are appropriate; i.e., they will generate the appropriate sequence — 0 1

Exercise 4. Placing parentheses. Let p_n be the number of ways to place parentheses.

- Author clearly displays a recurrence relation with initial conditions — 0 1 2
- Author gives a cogent argument for the recurrence — 0 1 2
- Author's recurrence is correct; it is $p_n = \sum_{i=1}^{n-1} = p_i p_{n-i}$, for $n \geq 2$ — 0 1 2
- The initial conditions given are correct: $p_0 = 0$ and $p_1 = 1$. — 0 1

Exercise 5. How many pizza slices can you make with n straight cuts with a pizza cutter such that each cut intersects all previous cuts and no three cuts intersect in a single point. Denote this number by R_n.

- Author clearly displays a recurrence relation with initial conditions — 0 1 2
- Author gives a cogent argument for the recurrence — 0 1 2
- Author's recurrence is correct; it is $R_n = R_{n-1} + n$, for $n \geq 1$ — 0 1 2
- The initial conditions given are correct: $R_0 = 1$. — 0 1

Exercise 6. The number of "words" with n "letters" from $\{\alpha, \beta, \gamma\}$ with no consecutive αs. Let w_n be this number.

- Author clearly displays a recurrence relation with initial conditions — 0 1 2
- Author gives a cogent argument for the recurrence — 0 1 2
- Author's recurrence is correct; it is $w_n = 2w_{n-1} + 2w_{n-2}$, for $n \geq 2$ — 0 1 2
- The initial conditions given are correct: $w_0 = 1$, $w_1 = 3$. — 0 1 2

Overall quality and presentation: lack of high quality will lose points, as will misused notation or other errors not captured in the objectives described in the above rubric (these deficiencies will be marked with a star on Author's paper). — 0 −1 −2 −3

Total ______/35.

[1] For each of the first objectives of this assignment, if the notation involved in the recurrence is not well-defined, then points will be lost; e.g., if the recurrence is $a_n = a_{n-1} + a_{n-2}$, then a_n must be clearly defined. Also, the independent variable n must be clearly defined.