THE USE OF BOUNDED GRADING SCALES IN HIGHER EDUCATION

P.H. Vossen[1], S. Ajit[2]

[1]*Edumetrics R&D (GERMANY)*
[2]*University of Northampton (UNITED KINGDOM)*

Abstract

Using bounded grading scales has a long tradition in higher education and on the job training. It is used in a variety of ways to assess what students or trainees have learned and how well they can apply their newly acquired knowledge to accomplish certain tasks or solve given problems. The assessment result may take the form of a written report by the assessor indicating how many tasks or problems the assessees have accomplished and why they failed on other ones. Using assessment protocols, including ordered grades on a bounded grading scale, can be challenging and inefficient to work with. For instance, assessors may find it difficult to decide which of two successive grades better represents a given assessment result, when the grading scale is rather coarse. In this methodologically oriented paper, the universal concept of a bounded grading scale is proposed as a solution. We will show how to calculate with grades on a bounded scale such that the sum of two grades is again a grade and the scalar product of a non-negative real number and a grade result in a rescaled or weighted grade. We will then show how sums of weighted grades play the same role on bounded scales as the more familiar concept of the weighted arithmetic average of unbounded real numbers. Furthermore, we will introduce the concept of a neutral grade as the threshold between "negative" and "positive" grades. Bounded grading scales together with the above-mentioned relations and operations may be used to solve several recurrent problems in educational assessment in the same framework. We will demonstrate it exemplarily for three scenarios: (1) A complex assignment shall be assessed using many criteria all using the same bounded scale; (2) as before, but now each criterion may be associated with a different bounded scale; (3) In Group-Peer Assessment, the assessor calculates a grade for the group product as well as grades for the respective student contributions based on peer assessments of the students (using the preceding scenario's); then, a scoring rule maps student contributions to student grades such that the mean student grade equals the group grade. In the end, we will discuss some methodological issues regarding the use of bounded scales in the field of educational assessment and conclude with suggestions for future work on how our basic framework can be further generalized to cover still more assessment styles like disjunctive and conjunctive assessment.

Keywords: Bounded scale, n-point scale, grading scale, sum of grades, rescaled grades, weighted mean grade, scale mappings, percentage scale, signed percentage scale, constrained percentage scale.

1 INTRODUCTION TO BOUNDED GRADING SCALES

Bounded grading scales have a long tradition in higher education, vocational instruction and on the job training. They are used in a variety of ways to assess what students or trainees have learned and how well they can apply their newly acquired knowledge to accomplish certain tasks or to solve particular problems. The assessment result may take the form of a written report by the assessor indicating how many tasks or problems the assessees have accomplished and why they failed on the other ones. Students or trainees can then use that information to prepare for a rehearsal of the same test or task, or for the next one. In the case of a final examination, assessors will use it to justify and report their judgments and decisions regarding the status and future positions of the assessees at the institution or workplace.

Using assessment protocols, including ordered grades on a grading scale, can be both challenging and inefficient to work with when it comes to aggregating the grades of a portfolio or sequence of assessments. Also, assessors find it often difficult to decide which of two successive grades better represents a given assessment result, when the grading scale in use is rather coarse, e.g., A up to G, or 1 to 10. This may also explain the invention of alphanumeric grading scales with special codes like B− and C+ (UK grading scale) to solve grading conflicts.

Such workarounds can't go on forever. In this paper, the generalized concept of a bounded grading scale between a lower bound (LB) and upper bound (UB) is proposed as an all-purpose solution applicable in a variety of ways. In other words: a grading scale no longer consists of distinct alphanumeric codes; instead, a grade may be any point on a line segment from LB to UB. By associating numbers with LB and UB, the grading scale will be denoted by $[LB, UB]$: the (ordered) range of real numbers between LB and UB. For example: $[0,1]$ denotes the standard (unsigned) percentage scale between 0 and 1; $[-1, +1]$ denotes the signed percentage scale between -1 and $+1$; $[-2, +2]$ denotes the bipolar Likert scale between -2 and $+2$; and $[1,10]$ denotes the 10-point grading scale between 1 and 10. Note that a bounded scale is not necessarily or implicitly restricted to the integer numbers (or their alphanumeric codes) on that scale, although it often is – for historical, pragmatic, or other reasons. In the latter cases, we will call the integers or codes the *anchors* of the underlying (continuous) scale. When working with bounded grades, it is tacitly assumed that the full bounded scale is meant, otherwise it would not be possible to calculate such a simple thing as the arithmetic average of a set of grades (if that would make sense at all, see below).

In previous papers [1-12] we have gradually worked out a rich and fascinating calculus of grading on bounded scales. It is based on the mathematical notion of a *module*, i.e., a system consisting of elements (here: grades) together with ordering relations (smaller, equal, larger) between those elements and two operations on them (sum and scalar product). In this paper, to avoid being accused of self-plagiarism, we will refrain from building up that calculus from scratch. Instead, we will assume that most readers of the current paper are more or less familiar with it. However, for those not yet familiar, we will (a) give a concise motivation for introducing bounded scale theory in educational assessment, (b) summarize its main principles, vocabulary, and notations, and (c) highlight the main results of our ongoing (and certainly not yet finished) deep research on this topic (see section 2). The current paper can thus be understood without the need to go back to our previous papers. For more general aspects related to the topics of this paper, see the reviews in [13-19].

Bounded scale systems may be used to solve several problems that emerge quite commonly in educational assessment and measurement. In section 3, we will demonstrate their use for three selected scenarios, the first two rather simple and frequently used, the third more complex and innovative:

(1) A complex assignment shall be assessed using many criteria all using the same bounded scale. The overall assessment may be reported as the bounded sum of weighted grades.

(2) Same as before, but now each criterion is associated with a possibly distinct bounded scale. The simplest solution will be to map the grades to a common bounded scale, e.g. the standard percentage scale, and to calculate the weighted sum of grades on that common scale.

(3) In the two preceding scenarios, the different grades play a symmetric role in the aggregation operation called "sum of weighted grades": the weighted grades can be added up in any order, it makes no difference at all. In Group-Peer Assessment's scoring rule, the situation is quite different: we have a group product grade on one scale and so-called students' process grades (to be called student contributions for mnemonic reasons) on another scale, and the scoring rule will map them to a third student grading scale such that the respective roles of group score and student contributions in the scoring rule cannot be exchanged at will.

Finally, in section 4, we will discuss some general issues regarding the emerging theory and methodology of bounded scales in education. Because this theoretical and methodological framework is still in its infancy, we will conclude with suggestions for future work on how the current assessment framework can be further extended, generalized, and unified to cover even more assessment and evaluation types with a minimum of structural changes, e.g., introducing more parameters.

2 BOUNDED SCALING THEORY AND METHODOLOGY

We will show how to calculate with grades on a bounded scale $[LB,UB]$ such that the sum of two grades is again a grade, and any grade may be multiplied by a positive decimal number, called its scalar weight, to obtain a weighted or rescaled grade. We will then show that sums of weighted grades play the same role on bounded scales as the more familiar concept of the weighted average of a set of real ("decimal") numbers. Furthermore, we will introduce the important concept of a neutral grade, which marks the threshold between "negative" and "positive" grades. For now, the neutral grade equals the midpoint between LB and UB; however, any grade may play this role (see section 4).

2.1 Bounded scales are paramount in educational assessment

In education, human evaluators, *aka* assessors, almost always rely on some sort of bounded scale to express their judgements about the quality of educational outcomes or performances of students. The bounded scales come in all sorts of disguises:

(1) A unipolar n-point scale exhibiting n discrete points in ascending or descending order.
(2) A bipolar $(2k + 1)$-point scale ranging from $-k$ over 0 to $+k$, in ascending order.
(3) The percentage scale, ranging from 0 to 1, usually denoting proportional achievement.
(4) The signed percentage scale, ranging from -1 to $+1$, usually denoting the degree of under- or overachievement relative to some sort of standard or default achievement (usually: 0).

Unfortunately, there is no universal agreement based on conclusive empirical evidence over which type or variant of bounded scale is the best one to use [18-19]. Instead, it is invariably a matter of convention bound by historical and cultural consensus-making over centuries in particular professional communities, which bounded scale, or scales are (compulsorily) used at a given institution in a given country. In practice, education appears to be relatively slow in taking up insights and methods of modern measurement theory, although there has been written and spoken a lot about it for more than a century and even though for many years there are already software tools on the market to assist in psychometric (e.g., cognitive) measurement, that are also applicable in educational contexts.

Furthermore, current educational assessment skills and practices are heavily influenced by a kind of statistical thinking that treats grading as an error-prone process in need of statistical techniques to minimize those errors by averaging them out. The idea that human grading may after all be relatively robust and is rather in need of simple and versatile models and methods to operate with large batches of grades on various grading scales in ways that can be fine-tuned to the context and purpose of educational assessment has not yet got wide acceptance and active support in academia, e.g., when teaching and training preservice teachers.

These are some of the observations and considerations that prompted us to take the notion of bounded scales seriously and to develop a theory and methodology of bounded scaling together with appropriate tools (spreadsheet templates using Excel or Google) for the educational practitioner (teachers, lecturers, trainers, etc.). Without such a theory and methodology, the risk is high, that educational research and practice is bound to use inferior methods and techniques of assessment without even being aware of it. Empirical research and case studies alone will not be very helpful in such a situation, because (a) they almost invariably (must) take existing grading practices for granted (on face value), (b) they will not be able to systematically detect and identify measurement flaws or errors in the data they are supposed to work with, and (c) therefore, they cannot deliver well-founded solutions for recurrent issues or problems and for newly emerging challenges (e.g., mass education, big data) awaiting constructive proposals.

2.2 Basic facts and concepts of bounded scales

The first and foremost assumption about achievement scales used in education is, that they are bounded. That is, it is always possible to define some lowest level of achievement and some highest level of achievement, given the field and goal of educational instruction (faculty/module/course). It doesn't imply, that those lower and upper levels of achievement (outcome or performance) will be context free and eternally fixed. It only means, that given a certain course, for instance, with well-defined goals and subject matters, and related achievement criteria (component dimensions), it will always be possible to propose a scale ranging from the lowest level to the highest level of achievement. Let we denote the lowest level by A, and the highest level by B. The domain of scale numbers is then $[A, B]$.

Implicit in the above formulation is the ranking of achievements from the worst to the best on the basis of observed quality criteria of outcomes and/or performances. We want to make this explicit by saying that the grading scale imposes the ordering relationship $\leq$, "smaller than or equal to", on its elements. If we assume, for the moment, that the levels, or degrees, of achievement are increasing (resp. decreasing) more or less uniformly and continuously from the lower bound A to the upper bound B, then it will also appear natural to place a kind of threshold level of achievement at the middle of the grade domain, i.e., at ½$(A + B)$. Grades below this middle level are called negative, grades above this level are called positive. The threshold level itself will be called the neutral grade and denoted by σ. The neutral grade plays a special role on the grading scale, so it will be part of its profile (see Fig. 1):

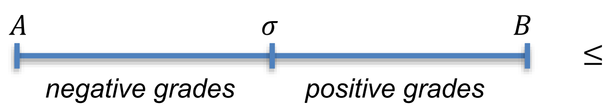

Figure 1. An ordered bounded scale with endpoints A and B and neutral grade σ in the middle.

Let us now deal with the basic operations on the grading scale. Given two grades x and y, we want to be able to add them in such a way that the result, called bounded sum, and denoted by $x \oplus y$, will also be a grade in $[A, B]$, i.e., $A \le x \oplus y \le B$. Generally, $x \oplus y$ can not be the same as $x + y$, because that would lead to a constraint on A and B that might turn out to be false (take e.g. the case where $x = y = A$ that can only satisfy $A \le x + y \le B$ if $2A \le B$; then take $A = 5$ and $B = 9$). Luckily, given $\sigma = ½(A + B)$, there exists a simple definition of $\oplus$, that also has got an intuitive graphical explanation [9]:

$$\frac{x \oplus y - A}{B - x \oplus y} \stackrel{\text{def}}{=} \frac{x - A}{B - x} \times \frac{y - A}{B - y} \quad \leftrightarrow \quad x \oplus y = A + (B - A) \times \frac{(x-A)(y-A)}{(x-A)(y-A)+(B-x)(B-y)} \tag{1}$$

Now we can precisely explain the special role played by the neutral grade. If we take $y = \sigma$, then it turns out that $x \oplus \sigma = x$, i.e., adding the neutral grade σ to any grade x, doesn't have any effect on the result. It plays the same role as 0 for the integers or for the decimal numbers.

Looking at (1), it may come as no surprise, that there is a related operation called bounded multiplication, and denoted by the operator symbol $\odot$, by which a grade x is pre-multiplied by a scalar factor $r \ge 0$:

$$\frac{r \odot x - A}{B - r \odot x} \stackrel{\text{def}}{=} \left(\frac{x - A}{B - x}\right)^r \quad \leftrightarrow \quad r \odot x = A + (B - A) \times \frac{(x - A)^r}{(x - A)^r + (B - x)^r} \tag{2}$$

Again, we have a special case for $r = 0$, as $0 \odot x = \sigma$. Also note that for $r = 1$ it follows that $1 \odot x$ equals x, just like the role played by the number 1 when multiplying integer or decimal numbers. Finally, for $r \to \infty$ we have: $r \odot x \to A$ when $x < \sigma$ and $r \odot x \to B$ when $x > \sigma$. More details can be found in our previous papers.

2.3 Advanced constructions on bounded scales

In the previous subsection, we have introduced the basic relations between, and operations on bounded grades. We are still missing three important constructions that go beyond those basics:

a) the sum of multiple grades;
b) the equally weighted average of multiple grades;
c) the weighted average of multiple grades.

With "multiple grades" we mean a n-tuple of grades indexed by subscript i: $x_1, \dots, x_n$. Without taking the trouble of deriving those constructions step-by-step from the definitions (1) and (2), we just list them here and refer the interested reader to our previous papers [1-12]:

a) The bounded sum of $x_1, \dots, x_n$:

$$\frac{\oplus_{i=1}^n x_i - A}{B - \oplus_{i=1}^n x_i} \stackrel{\text{def}}{=} \prod_{i=1}^n \frac{x_i - A}{B - x_i} \quad \leftrightarrow \quad \oplus_{i=1}^n x_i = A + (B - A) \times \frac{\prod_{i=1}^n (x_i - A)}{\prod_{i=1}^n (x_i - A) + \prod_{i=1}^n (B - x_i)} \tag{3}$$

b) The equally weighted (sometimes confusingly called "unweighted") average of $x_1, \dots, x_n$:

$$\frac{\frac{1}{n} \odot (\oplus_{i=1}^n x_i) - A}{B - \frac{1}{n} \odot (\oplus_{i=1}^n x_i)} \stackrel{\text{def}}{=} \sqrt[n]{\prod_{i=1}^n \frac{x_i - A}{B - x_i}} \quad \leftrightarrow \quad \frac{1}{n} \odot (\oplus_{i=1}^n x_i) = A + (B - A) \times \frac{\sqrt[n]{\prod_{i=1}^n (x_i - A)}}{\sqrt[n]{\prod_{i=1}^n (x_i - A)} + \sqrt[n]{\prod_{i=1}^n (B - x_i)}} \tag{4}$$

c) The weighted average of $x_1, \dots, x_n$ with potentially different weights $w_1, \dots, w_n$ summing up to 1:

$$\frac{\oplus_{i=1}^n w_i \odot x_i - A}{B - \oplus_{i=1}^n w_i \odot x_i} \stackrel{\text{def}}{=} \prod_{i=1}^n \left(\frac{x_i - A}{B - x_i}\right)^{w_i} \quad \leftrightarrow \quad \oplus_{i=1}^n w_i \odot x_i = A + (B - A) \times \frac{\prod_{i=1}^n (x_i - A)^{w_i}}{\prod_{i=1}^n (x_i - A)^{w_i} + \prod_{i=1}^n (B - x_i)^{w_i}} \tag{5}$$

In some situations, it may be necessary to invoke other kinds of constructions, that often involve some sort of pre-processing or standardization of the original grades. For example, we have found the following conversion of grades very useful in the case that the assessor wants to exclude the extreme grades A and B, that happen to have a drastic effect in the above formulae: a single grade of A or B forces all sums and averages to become A or B, resp., irrespective of the other grades entering the formulae. This can be avoided without losing or corrupting the original data by choosing a small delta, e.g. $\delta = 0.001$, and applying the following conversion (indicated by the asterisk *) to all grades $x_1, \dots, x_n$ (including the culprits, of course):

$$x_i^* \stackrel{\text{def}}{=} x_i + \delta \times \left(\frac{A + B - 2x_i}{B - A}\right) \tag{6}$$

It can be easily proved that the absolute difference between the original grade x_i and the converted grade x_i^* will not exceed δ (hardly noticeable) and that the grades A and B will have been replaced by $A + \delta$ and $B - \delta$, resp. The only grade that will not be replaced is the neutral grade $\sigma = ½(A + B)$.

In a more general setting, one may have reason to apply a particular transformation to the original data before invoking one of the above aggregation formulae (3-5). Very often this will be the case if the original grades are coming from one bounded scale, and the calculated average of the grades shall be reported on another bounded scale. This is exactly what we will do in the next section, scenario III, Group-Peer Assessment!

3 EDUCATIONAL APPLICATIONS OF BOUNDED SCALES

In the foregoing section, we have given a concise overview of bounded scales in the abstract. As promised in the beginning of this paper, we want to apply this theory and methodology to three typical scenarios in the field of educational assessment. For simplicity, we will be using dummy data.

3.1 Scenario I: Aggregation of grades on the percentage scale.

Assume you have taught a course on creative writing in English and all students must deliver an essay of at least 8 pages no later than 4 weeks after the formal end of your lectures. You have shared with the students a list of 5 criteria or rubrics (e.g., overall structure, readability, grammar, citing and referencing, etc.) that you will use for assessing and grading each essay. Thus, we are talking about an individual assignment (no group work) and criteria-based assessment (not holistic, i.e., one-shot, assessment). All criteria will be assessed on the standard percentage scale $[0,1]$. You have ordered the criteria from most important to least important, with the relative weights of 30%, 25%, 20%, 15%, and 10%. Thus, the overall assessment will be reported as the sum of weighted grades:

TABLE I. ESSAY WRITING, PERCENTAGE SCALE, WEIGHTED SUM ON THE PERCENTAGE SCALE

Criterion	Weight w_i	Grade x_i	Weighted Grade
I.1	$.30$	$.70$	$.30 \odot .70 = .56$
I.2	$.25$	$.40$	$.25 \odot .40 = .47$
I.3	$.20$	$.75$	$.20 \odot .75 = .55$
I.4	$.15$	$.75$	$.15 \odot .75 = .54$
I.5	$.10$	$.60$	$.10 \odot .60 = .51$
	1.00	final grade: $\bigoplus_{i=1}^{5} w_i \odot x_i = .64$	

Because the criteria are already scored on the percentage scale, you do not need to apply some scale-to-scale conversion in order to compute the weighted criteria grades $w_i \odot x_i$: they can be calculated directly using formula (2) substituting $A = 0$ and $B = 1$. For the final grade you will apply formula (5) with different weights for the 5 criteria. Whether you use the right side of formula (5) or the left side of it is entirely up to you; the left side is somewhat easier to express in words: calculate the geometric mean GM of the simple transformation $\frac{x}{1-x}$ of the individual grades x and then compute the reverse transformation $\frac{GM}{1+GM}$. That's all.

3.2 Scenario II: Aggregation of grades on different n-point scales.

Now assume you have taught a course on Database Design as part of a larger module DBMS. The students must deliver a DB model specification for a local library based on your semi-formal description in plain language. The specification shall include a list of all proposed attributes (typed fields with further properties) as well as an Entity-Relationship Diagram in standard normalized form. This time, the students will work in self-chosen teams of 3 to 5 students. They have got 6 weeks to deliver it. As part of the assignment, you have selected and shared a set of 5 DB quality criteria for relational databases (the subject of your course). Again, we are talking here about criteria-based assessment. Although the semi-formal description of the intended database will be explicit in its requirements of what to cover or not, the teams will still have a rather high degree of freedom how to transform that in an adequate database design: thus, it is an open-ended assignment, i.e., there are several equally valid possible solutions for this design exercise. Because the quality criteria are rather diverse in how they can be assessed, they will be associated with different n-point grading scales. So, in Table II, we have added two columns, one for the lower bound and one for the upper bound of the scale. However, their importance will be roughly equal, so you have assigned equal weights of 20%, and the overall assessment will be reported as the sum of unweighted grades.

TABLE II. DATABASE DESIGN, N-POINT SCALES, UNWEIGHTED SUM ON THE PERCENTAGE SCALE

Criterion	Lower Bound	Upper Bound	Weight w_i	Grade x_i	Weighted Grade
II.1	-1	$+1$	$.20$	0	$\frac{1}{5} \odot .70 = .54$
II.2	1	5	$.20$	4	$\frac{1}{5} \odot .80 = .57$
II.3	1	5	$.20$	2	$\frac{1}{5} \odot .40 = .48$
II.4	1	3	$.20$	3	$\frac{1}{5} \odot .99 = .71$
II.5	-2	$+2$	$.20$	0	$\frac{1}{5} \odot .50 = .50$
			1.00	final grade: $\frac{1}{5} \odot (\bigoplus_{i=1}^{5} x_i) = .78$	

Note that the grades x_i have been converted to the percentage scale using a simple conversion rule. Note also, that for criterion II.4, you must apply formula (6) to avoid issues with grade $x_4 = 3$ on its scale from 1 to 3. To calculate the final grade, you will use formula (5), which is evidently somewhat simpler to compute than formula (6) because you have only once to apply scalar multiplication by $\frac{1}{5}$.

3.3 Scenario III: Aggregation of grades using a non-associative scoring rule

In the two preceding scenarios, the grades are assumed to play a symmetric role in the aggregation called “sum of weighted grades”: the weighted grades can be summed together in any order, it makes no difference at all. In Group-Peer Assessment, the situation is quite different, at least in the final stage: we will have a group score on one scale as well as a set of student contributions on another scale, and a scoring rule that merges them together onto a third scale such that the respective roles of group score and student contributions cannot be exchanged at will.

In order to understand and appreciate the situation in its full complexity we must be careful in our terminology and notation. First, let us denote by t the group product score, which is nothing else but a grade on the percentage scale for the group’s product or outcome using whatever product quality criteria you deem adequate for the group assessment. Thus, this is just an instance of Scenario I we have described above.

Secondly, you want the students in a group to co-assess each other on a set of group dynamics criteria that you provided and shared with the group. The reason that you “outsource” the assessment of the individual members of the group with respect to their contribution to the group work eventually leading to their joint product is, that you haven’t really a deep insight of what is going on between the peers.

In order to facilitate peer assessment of the group’s processes without unduly putting extra workload on the students you will select a small set of observable process criteria that can be easily identified and assessed by the students using distinct N-point grading scales. Let us denote the n students in a group by i in their role as assessee (being assessed) and by j in their role as assessor (assessing). Then we may denote by r_{ij} the peer rating for student i given by his/her peer j, converted to a percentage (always possible). Thus, this is an instance of Scenario II we have described above.

Then, each student i receives $n-1$ peer ratings r_{ij} for $j = 1, \ldots, i-1, i+1, \ldots n$ which will be aggregated (using formula (5) with $A = 0$ and $B = 1$) to a student's rating r_i, taking possible peer student weightings w_j into account:

$$\frac{r_i}{1-r_i} \stackrel{\text{def}}{=} \prod_{j=1, j\neq i}^{n} \left(\frac{r_{ij}}{1-r_{ij}}\right)^{w_j} \tag{7}$$

To calculate the mean student rating $\bar{r}$ (*aka* group rating) we repeat aggregating once more over the students i with the same weights w_i:

$$\frac{\bar{r}}{1-\bar{r}} \stackrel{\text{def}}{=} \prod_{i=1}^{\mathrm{n}} \left(\frac{r_i}{1-r_i}\right)^{w_i} \tag{8}$$

Now we can define the student contribution c_i on the signed percentage scale $[-1, +1]$ in terms of the student rating r_i, the mean student rating $\bar{r}$, and a system parameter $p \geq 0$, called the peer impact, by means of the following model equation:

$$\frac{1+c_i}{1-c_i} \stackrel{\text{def}}{=} \left(\frac{r_i}{1-r_i} \div \frac{\bar{r}}{1-\bar{r}}\right)^{p} \tag{9}$$

The peer impact enables us to adjust the influence of the relative student rating (the quotient inside) on the student contribution, and thus indirectly on the final student score (to be defined in equation (10)).

This student contribution c_i has the following interesting and plausible properties in terms of student ratings. First, if r_i equals $\bar{r}$, then $c_i = 0$, the neutral contribution. Second, if r_i equals 0, then $c_i = -1$, the smallest contribution. Third, if r_i equals 1, then $c_i = 1$, the largest contribution. Fourth, if $p = 0$, then c_i equals the neutral contribution, and if p grows larger and larger, then c_i approaches either -1 or $+1$ depending on the location of the student rating relative to the mean student rating. Thus, the behaviour of the student contribution appears to reflect the behaviour of the student ratings very well and in a standardized manner.

We are now in the best position to propose the GPA scoring rule combining the group grade t, another system parameter z between 0 and 1, and the student contribution c_i:

$$s_i \stackrel{\text{def}}{=} t^{z^{c_i}} \tag{10}$$

where s_i stands for the final student grade on the constrained percentage scale between $\sqrt[z]{t}$ and t^z. The domain of the constrained percentage scale can be anything between the full percentage scale $[0,1]$ in case of $z = 0$ (no constraint), and the one-grade-set $\{t\}$ in case of $z = 1$ (full constraint). In the following example calculation of the final student grades we have assumed a (default) peer impact p of 1, a group grade t of 0.70 (70%), and a (default) spread constraint z of ½, so that the final grades lie on a constrained percentage scale between t^2 and $\sqrt{t}$:

TABLE III. STUDENT RATINGS, CONTRIBUTIONS AND GRADES ON A CONSTRAINED PERCENTAGE SCALE

Students	Ratings	Contributions	Grades
1	$.55$	$-.21$	$.66$
2	$.60$	$-.11$	$.68$
3	$.65$	$-.01$	$.70$
4	$.70$	$.11$	$.72$
5	$.75$	$.23$	$.74$
bounded mean	$\mathbf{.65}$	$\mathbf{.00}$	$\mathbf{.70}$

Note the interesting property that $s_i = t$ if $c_i = 0$ (appearing in the table as -0.01, due to rounding errors), thus a neutral student contribution of 0 leads to a student grade of t, i.e., the group grade, which is neither positive nor negative. What may be more surprising: the mean student grade according to equation (10) equals the group grade t (because the mean student contribution equals 0, as is easy to verify by equation (9), and then we are back to the previous observation).

4 DISCUSSION AND CONCLUSIONS

The three scenarios in the previous section exemplify the broad spectrum of applications covered by the theory and methodology of bounded scales. Starting with some simple assumptions and constructions in section 2, we have arrived at a complete solution of the Group-Peer Assessment problem for which there have been only incomplete, mostly inadequate solutions in the literature, e.g., [20, 21]. Looking again at the construction in Scenario III, it will be clear that no empirical research would have been enough to arrive at the solution proposed here, i.e., formulae (7) up to (10). Indeed, empirical research has its inherent limitations, weak points, and blind spots, by focussing primarily on collecting, analysing, and interpreting surface data in a statistically sound way, but not putting enough attention to and resources on studying the deep principles and regularities that underly and generate those surface data in the first place.

Once we have a theoretically sound solution, researchers may go ahead and find out what the limitations (from the perspective of empirical research) of this particular solution are, and what would be required from theoretical and methodological research to overcome those limitations. On the other hand, it would be interesting to review past empirical research on the statistical properties of peer assessment models and data in order to find out to which extent its conclusions or recommendations might have been distorted or biased by the fact that the laws of bounded scales were ignored and hence, that researchers might have been using inappropriate data analytics.

In the meantime, we are aware of the fact, that the theory and methodology of bounded scales is still in its infancy and needs further investigation and deep analysis. One of the analytical research directions we are currently working on is the possibility of generalizing our framework to cover disjunctive and conjunctive assessment models, because we know by informal communication that some assessors are using such models. In contrast to the balanced aggregation model we have proposed here (the neutral grade lies exactly in the middle of the bounded scale), in the generalized disjunctive and conjunctive models the neutral grade may be shifted to either side of the middle grade. In two extreme cases, the neutral grade will even coincide with either the lower bound ($\sigma = A$) or the upper bound ($\sigma = B$) of the adopted scale. To cover such extensions of our basic framework, it will be necessary to replace the initial definitions of bounded sum and bounded product.

REFERENCES

[1] P. H. Vossen, “A truly generic performance assessment scoring system (PASS),” in *INTED2011 Proceedings*, pp. 2448-2458, IATED, 2011.

[2] P. H. Vossen, “Educational Assessment Engineering: A Pattern Approach,” in *Soft Computing Applications*, edited by Valentina Emilia Balas, Lakhmi C. Jain, and Branko Kovačević, 605–19. Cham: Springer International Publishing, 2015.

[3] I. G. Kennedy and P. H. Vossen, “Software engineering teamwork assessment rubrics: combining process and product scoring,” *Assessment in Higher Education*, Manchester, 2017

[4] P. H. Vossen, “Distributive Fairness in Educational Assessment,” in *Soft Computing Applications*, edited by Valentina Emilia Balas, Lakhmi C. Jain, and Marius Mircea Balas, 381–94. Cham: Springer International Publishing, 2018.

[5] P. H. Vossen, S. Ajit and A. Dean, “Quasi-Arithmetic Scoring and Group Project Marking,” in *ICERI2019 Proceedings*, 8395–8405. Seville, Spain: IATED, 2019, Retrieved from https://doi.org/10.21125/iceri.2019.2002.

[6] P. H. Vossen,“Scoring Models for Peer Assessment in Team-Based Learning Projects,” in *EDULEARN19 Proceedings*, 9704–13. Palma, Spain: IATED, 2019, Retrieved from https://doi.org/10.21125/edulearn.2019.2424.

[7] P. H. Vossen and S. Ajit, “Towards a Paradigm Change in Group and Peer Assessment in Software Engineering Education,” in *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T)*, 1–10, 2020. Retrieved from https://doi.org/10.1109/CSEET49119.2020.9206197.

[8] P. H. Vossen and S. Ajit, “Fuzzy Scoring Theory Applied to Team-Peer Assessment ~ Additive vs. Multiplicative Scoring Models on the Signed or Unsigned Unit Interval,” in *Soft Computing Applications*, edited by Valentina Emilia Balas, Lakhmi C. Jain, Marius Mircea Balas, and Shahnaz N. Shahbazova, 84–111. Cham: Springer International Publishing, 2021.

[9] P.H. Vossen and S. Ajit, “The Challenge of Peer Assessment in Higher Education,” in *IC-ERI2022 Proceedings*, 8603–12. Seville, Spain: IATED, 2022. Retrieved from https://doi.org/10.21125/iceri.2022.2270.

[10] P. H. Vossen and S. Ajit, “Bounded Scales ~ A New Type of Scale for Educational Practice and Research.” In *EDULEARN23 Proceedings*, 6912–22. Palma, Spain: IATED. 2023. Retrieved from https://doi.org/10.21125/edulearn.2023.1818.

[11] P. H. Vossen and S. Ajit, “Group-Peer Assessment from a Quantitative Point of View,” in EDU-LEARN23 Proceedings, 6979–89. Palma, Spain: IATED. 2023. Retrieved from https://doi.org/10.21125/edulearn.2023.1831.

[12] P. H. Vossen and S. Ajit, “Assessment and Measurement in Higher Education with Multiple Criteria, Scales, and Raters ~ A Theoretical Approach and Practical and a Practical Example.” In *INTED2024 Proceedings*, 7469–75. Valencia, Spain: IATED, 2024. Retrieved from https://doi.org/10.21125/inted.2024.1963.

[13] M. Alqassab, J-W. Strijbos, E. Panadero, J. Fernández Ruiz, M. Warrens, and J. To. “A Systematic Review of Peer Assessment Design Elements,” *Educational Psychology Review* 35 (1): 18, 2023. Retrieved from https://doi.org/10.1007/s10648-023-09723-7.

[14] F. Blanco-Mesa, E. León-Castro, and J. M. Merigó, “A Bibliometric Analysis of Aggregation Operators,” *Applied Soft Computing 81 (August): 105488*, 2019. Retrieved from: https://doi.org/10.1016 /j.asoc.2019.105488.

[15] G. T. L. Brown, “The Past, Present and Future of Educational Assessment: A Transdisciplinary Perspective,” *Frontiers in Education 7* (November), 2022. Retrieved from https://doi.org/10.3389/feduc.2022.1060633.

[16] M. Fertalj, L. J. Brkić, and I. Mekterović, “A Systematic Review of Peer Assessment Approaches to Evaluation of Open-Ended Student Assignments,” In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1076–81, 2022. Retrieved from https://doi.org/10.23919/MIPRO55190.2022.9803770.

[17] L. R. Martínez, D. Luis, and J. C. Acosta, “Aggregation Operators Review-Mathematical Properties and Behavioral Measures,” *International Journal of Intelligent Systems and Applications*, Vol. 7, No. 10, p. 63-76, 2015. Retrieved from http://repositorio.unne.edu.ar/xmlui/handle/123456789/29006.

[18] G. Norman, “Likert Scales, Levels of Measurement and the ‘Laws’ of Statistics,” *Advances in Health Sciences Education*, Vol 15, No 5. p. 625–32, 2010. Retrieved from https://doi.org/10.1007/s10459-010-9222-y.

[19] H. Taherdoost, “What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale,” *SSRN Scholarly Paper*, Rochester, NY, 2019. Retrieved from: https://papers.ssrn.com/abstract=3588604.

[20] S. Sharp, “Deriving individual student marks from a tutor’s assessment of group work,” *Assessment & Evaluation in Higher Education*, Vol. 31, No. 3. p. 329-343, 2006.

[21] K. P. Nepal, “An approach to assign individual marks from a team mark: the case of Australian grading system at universities,” *Assessment & Evaluation in Higher Education*, Vol. 37, No. 5, p. 555-562, 2012.