DIVERSITY IS THE OPTIMAL EDUCATION STRATEGY: A MATHEMATICAL PROOF

OLGA KOSHELEVA
Department of Teacher Education
University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu

ABSTRACT. To enhance learning, it is desirable to also let students learn from each other, e.g., by working in groups. It is known that such groupwork can improve learning, but the effect strongly depends on how we divide students into groups. In this paper, we describe how to optimally divide students into groups so as to optimize the resulting learning. It turns out that the largest gain is attained when each of the resulting groups is a representative sample for the student population as a whole – i.e., when we have diversity.

Keywords: 1-D Model; Education Strategy; Dynamic Groups; General Multi-D

1. Introduction. Groupwork as A Way to Teach Better. Traditionally, students mostly learn from their instructor. The instructor presents the new material, asks the students to solve some related problems, and then provides individual feedback to students-explaining, to each student, his or her possible misunderstandings. Such an individual feedback is extremely helpful to the student. However, providing such an individual feedback requires a lot of time-especially in a class of reasonable size. So, if this feedback only comes from the instructor (and a Teaching Assistant), the amount of such feedback is limited. Moreover, in such situations, a significant amount of time is needed to grade the assignments of the whole class, so there is a significant delay between the test time and the time when students get their feedback.

It is well known that we can increase the amount of feedback – and decrease the delay of producing this feedback – if we also ask students from the class to provide useful feedback to each other. Different students have somewhat different misconceptions, so when a small group of students starts solving a problem together, they can often see each other’s mistakes and provide corrections – thus, teaching each other.

1.1. Groupwork is not a panacea. While in principle, groupwork is efficient; its efficiency depends on how we divide students into groups. If we simply allow students to group themselves together, often, strong students team together and weak students team together. Strong students already know the material, so they do not benefit from working together. Similarly, weak students are equally lost, so having them solve a problem together does not help; see, e.g., [1, 2, 3, 4, 5, 6, 9, 10, 11, 12].
1.2. How to divide students into groups? Since the efficiency of groupwork depends on the subdivision into groups, to make groupwork as efficient as possible, it is desirable to find the optimal way to divide students into groups. This is the problem that we study in this paper.

Comment. Several preliminary results – that eventually led to this paper – first appeared in (O. Kosheleva and V. Kreinovich, 2012).


2.1. Need for an approximate description. A realistic description of student interaction requires that we take into account a multi-D learning profile of each student: how much the students know of each part of the material, what is the student’s learning style, etc. Such a description is difficult to formulate and even more difficulty to optimize. Because of this difficulty, in this paper, we consider a simplified description of student interaction. Already for this simplified description, the corresponding optimization problem is non-trivial—but we succeed in solving it under reasonable assumptions.

2.2. How to describe the current state of learning. In this section, following (Kosheleva and Kreinovich, 2012), we consider a simplified (first approximation) model of student interaction. Our main simplifying assumption is that for each student, the degree to which this student learned the material can be characterized by a single number — crudely speaking, this student’s grade so far. In the following text, we will denote the number of students in the class by \( n \), and we will denote the degree of knowledge of the \( i \)-th student by \( d_i = 1, \ldots, n \). So, we arrive at the following definition.

**Definition 1.** Let \( n \) be an integer; we will call this integer a number of students. By a state of knowledge, we mean a tuple \( d = (d_1, \ldots, d_n) \) consisting of \( n \) nonnegative numbers.

2.3. Subdivision into groups. The following describes the general subdivision into groups.

**Definition 2.** Let \( n \) be a number of students. By a subdivision into groups, we mean a subdivision of the set \( \{1, \ldots, n\} \) into a finite number of non-intersecting subsets \( G_1, \ldots, G_m \) for which \( G_k \cap G_l = \emptyset \) for \( k \neq l \) and \( \bigcup_{k=1}^{m} G_k = \{1, \ldots, n\} \).

In this paper, we will mostly consider subdivision into groups of equal size.

2.4. How group work helps students: a description. If two students with degrees \( d_i < d_j \) work together, then the degree of knowledge of the \( i \)-th student increases. As we have mentioned earlier, if two students are at the same level of knowledge, there is not much that they can learn from each other. The more the \( j \)-th student knows that the \( i \)-th student doesn’t, the more the \( i \)-th student will learn. So, it is reasonable to assume that the amount of material that the \( i \)-th student learns is proportional to the difference \( d_j - d_i \), with some known coefficient of proportionality \( \alpha \). Thus, after the groupwork, the new level of knowledge of the \( i \)-th student is equal to \( d_i' = d_i + \alpha \cdot (d_j - d_i) \).
If more than two students work together, then each student learns from all the students from the group who have a higher degree of knowledge. For example, if three students, with original degrees of knowledge $d_i < d_j < d_k$, work together, then after the groupwork, their new levels of knowledge are equal to $d_i' = d_i + \alpha \cdot (d_j - d_i) + (d_k - d_i)$, $d_j' = d_j + \alpha \cdot (d_k - d_i)$, and $d_k' = d_k$. In general, we arrive at the following definition.

**Definition 3.** Let $n$ be the number of students, and let $\alpha > 0$ be a real number. For each state of knowledge $d = (d_1, \ldots, d_n)$ and for each subdivision into groups $G_1, \ldots, G_m$, the resulting state of knowledge $d' = (d_1', \ldots, d_n')$ is defined as follows: for every $k = 1, \ldots, m$ and for every $i \in G_k$, we have

$$d_i' = d_i + \alpha \cdot \sum_{j \in G_k, d_j > d_i} (d_j - d_i).$$

**2.5. Dynamic groups.** Subdivision into groups varies. Our objective is to find the subdivision which, at this moment of time, leads to the best gain. From this viewpoint, we need to find groups that work for a forthcoming short period of time, during which the change in grades – proportional to the coefficient $\alpha$ – is small. After this brief interaction, we can again gauge the student’s knowledge and, if needed, change the subdivision into groups to reflect what students learned. From this viewpoint, it is sufficient to consider small positive values $\alpha$.

**Comment.** Ideally, we should also take into account that there is a cost of group-changing: students spend some effort adjusting to their new groups.

**2.6. A reasonable objective function.** Our goal is to find a subdivision into groups for which the overall degree of knowledge is optimal. This optimal subdivision depends on how we gauge the overall degree of knowledge. In this paper, we consider the average grade $a \equiv \frac{1}{n} \cdot \sum_{i=1}^{n} d_i$.

**2.7. Results.** Let us consider the situation when we have $n = g \cdot m$ students, we know their degree of knowledge $d_1, \ldots, d_n$, and we want to subdivide these students into $m$ subgroups of $g$ students so as to maximize each of the three objective functions. Let us start with the simplest case $g = 2$, when we divide students into pairs.

**Proposition 1.** To maximize the average grade $a$, we divide the students into pairs as follows:

1. we sort the students by their knowledge, so that $d_1 \leq d_2 \leq \cdots \leq d_n$;
2. in each pair, we match one student from the lower half
   $$L_0 \equiv \{d_1, d_2, \ldots, d_{n/2}\}$$
   with one student from the upper half
   $$L_1 \equiv \{d_{(n/2)+1}, \ldots, d_n\}.$$

**Comment.** For reader’s convenience, all the proofs are placed in the special Appendix.
A similar result holds for groups of general size \( g \geq 2 \):

**Proposition 2.** For every \( g \geq 2 \), to maximize the average grade \( a \), we divide the students into groups as follows:

1. we sort the students by their knowledge, so that:
   \[ d_1 \leq d_2 \leq \cdots \leq d_n; \]
2. based on this sorting, divide the students into \( g \) sets:
   \[ L_0 = \{d_1, d_2, \cdots, d_{n/g}\}; \]
   \[ L_k = \{d_{k(n/g)+1}, \cdots, d_{(k+1)(n/g)}\}; \]
   \[ L_{g-1} = \{d_{(g-1)(n/g)+1}, \cdots, d_n\}; \]
3. in each group, we pick one student from each of \( g \) sets \( L_0, L_1, \cdots, L_{g-1} \).

**Interpretation.** In the optimal subdivision, each group should have, on average:

1. the same proportion of best prepared students as the total students population,
2. the same proportion of second-best prepared students as the total students population,
3. the same proportion of the least prepared students as the total students population.
4. In other words, each group should be representative of the student population as a whole. This representativeness is what is usually understood by diversity. Thus, an optimal group should be diverse.

**3. 1-D Model: A More Nuanced Description.**

**3.1. Main idea.** In the above analysis, we used a simplified model in which only the weaker students, with \( d_i < d_j \), benefit from the groupwork. In reality, stronger students, with \( d_j > d_i \), benefit too: when they explain the material to the weaker students, they reinforce their knowledge, and they may see the gaps in their knowledge that they did not see earlier. The larger the difference \( d_j - d_i \), the more the stronger student needs to explain and thus, the more this stronger student reinforces his or her knowledge. It is therefore reasonable to assume that the resulting increase in knowledge is proportional to the difference \( d_j - d_i \), with a different coefficient \( \beta > 0 \). Thus, we arrive at the following definition:

**Definition 4.** Let \( n \) be the number of students, and let \( \alpha > 0 \) and \( \beta > 0 \) be real numbers. For each state of knowledge \( d = (d_1, \cdots, d_n) \) and for each subdivision into groups \( G_1, \cdots, G_m \), the resulting state of knowledge \( d' = (d'_1, \cdots, d'_n) \) is defined as follows: for every \( k = 1, \cdots, m \) and for ever \( i \in G_k \), we have

\[
d'_i = d_i + \alpha \cdot \sum_{j \in G_k, d_j > d_i} (d_j - d_i) + \beta \cdot \sum_{j \in G_k, d_l > d_i} (d_l - d_i).
\]

It turns out that if we maximize either the average grade or the worst grade, then the optimal subdivisions are exactly the same as for the previously used (less nuanced) model:

**Proposition 3.** In the model described by Definition 4, to maximize the average grade \( a \), we divide the students into pairs as follows:
1. we sort the students by their knowledge, so that
\[ d_1 \leq d_2 \leq \cdots \leq d_n; \]
2. in each pair, we match one student from the lower half
\[ L_0 \equiv \{d_1, d_2, \cdots, d_{n/2}\} \]
with one student from the upper half
\[ L_1 \equiv \{d_{(n/2)+1}, \cdots, d_n\} \]

Proposition 4. In the model described by Definition 4, for every \( g \geq 2 \), to maximize the average grade \( a \), we divide the students into groups as follows:
1. we sort the students by their knowledge, so that
\[ d_1 \leq d_2 \leq \cdots \leq d_n; \]
2. based on this sorting, divide the students into \( g \) sets:
\[ L_0 = \{d_1, d_2, \cdots, d_{n/g}\}; \]
\[ L_k = \{d_{k-(n/g)+1}, \cdots, d_{(k+1)-(n/g)}\}; \]
\[ L_{g-1} = \{d_{(g-1)-(n/g)+1}, \cdots, d_n\}; \]
3. in each group, we pick one student from each of \( g \) sets \( L_0, L_1, \cdots, L_{g-1} \).

3.2. Discussion. In this more nuanced model, still the optimal solution is to have representative groups – i.e., has diversity.

4. From A Simplified 1-D Description to the General Multi-D Case. In the previous sections, we considered the simplified 1-D case, when the state of knowledge by a student \( i \) can be characterized by a single parameter \( d_i \). In general, the state of knowledge of a student \( i \) is characterized by several parameters: e.g., by degrees \( d_i^{(1)}, d_i^{(2)}, \) etc. to which this student progressed in each part of the course. When students study together in a group, they exchange their knowledge with respect to each part of the course. As a result, the new values \( d_i^{(k)'} \) are described by the formulas:
\[ d_i^{(k)'} = d_i^{(k)'} + \alpha \cdot \sum_{j \in G_k, d_j^{(k)} > d_i^{(k)}} (d_j^{(k)} - d_i^{(k)}) + \beta \cdot \sum_{j \in G_k, d_j^{(k)} > d_i^{(k)}} (d_j^{(k)} - d_i^{(k)}). \]

We want to maximize the average knowledge over all the parts of the class. For each part \( k \), the optimal average class grade is attained when we provide diversity over the values of the corresponding parameter \( d^{(k)} \), i.e., when each group is a representative sample with respect to this parameter \( d^{(k)} \). Thus, to make sure that our subdivision into groups is optimal with respect to all these parameters, we must make sure that each of these groups is a representative sample for all these parameters. No matter what parameter we choose, each group should have approximately the same distribution of this parameter as the student population as a whole.

In other words, in an optimal subdivision into groups, each group should be a representative sample of the student population as a whole. In short, we have proved that diversity is the optimal education strategy.
5. Conclusion. To enhance learning, it is desirable to also let students learn from each other by working in groups. It is known that such groupwork can improve learning, but the effect strongly depends on how we divide students into groups. We have proven that the largest gain is attained when each of the resulting groups is a representative sample for the student population as a whole – i.e., when we have diversity.

REFERENCES


Proofs

Proof of Proposition 1.
1. First, we note that maximizing the average grade is equivalent to maximizing the sum \( n \cdot a = \sum_{i=1}^{n} g_i' \) of the new grades, which is, in turn, equivalent to maximizing the overall gain \( \sum_{i=1}^{n} g_i' - \sum_{i=1}^{n} g_i = \sum_{i=1}^{n} (g_i' - g_i) \).
2. Let us take the optimal subdivision, and show that it has the form described in the formulation of Proposition 1.

   Indeed, in each pair, with degrees \( d_i \leq d_j \), we have a weaker student \( i \) and a stronger student \( j \). Let us prove that the optimal subdivision into groups, each stronger student is stronger (or of the same strength) than each weaker student. In other words, if we have two pairs \( d_i \leq d_j \) and \( d_i' \leq d_j' \), then \( d_i \leq d_j \), we will prove this by contradiction. Let us assume that \( d_i > d_j \). Let us then swap the \( i \)-th and the \( j \)-th students, i.e., instead of the original pairs \((i, j)\) and \((i', j')\), let us consider two new pairs \((i', j)\) and \((i, j')\). The corresponding two terms in the overall gain are changed from \( \alpha \cdot (d_i + d_j' - d_i - d_j) \) to \( \alpha \cdot (d_i - d_j' + d_i - d_j') \). The difference between the two expressions is equal to \( 2\alpha \cdot (d_i - d_j') \). Since we assumed that \( d_i > d_j' \), this difference is positive, which means that the above swap increases the overall gain. The possibility of such an increase contradicts to the fact that we have selected the subdivision for which the overall gain is already the largest possible. This contradiction shows that our assumption \( d_i > d_j \) is wrong, and thus, \( d_i \leq d_j \).

   Since every weaker-of-pair student is weaker than every stronger-of-pair student, all weaker-of-pair students from the bottom of the ordering of the degrees \( d_i \), while all the stronger-of-pair students form the top of this ordering – exactly as the formulation of Proposition 1 suggests.

3. To complete the proof, we need to prove that every subdivision satisfying the condition of Proposition 1 leads to the optimal average grade. Indeed, we know that one optimal subdivision satisfies this condition. One can check that for each such subdivision, the overall gain is equal to \( \sum_{i \in L_1} d_i - \sum_{j \in L_0} d_j \), where \( L_1 \) is the set of all the indices \( i \) from the upper half, and \( L_0 \) is the set of all the indices from the lower half. Thus, the overall gain for all such subdivisions is the same – and it is therefore exactly equal to the gain corresponding to the optimal subdivision. So, all subdivisions satisfying the condition of Proposition 1 indeed lead to the optimal average grade.

The proposition is proven.

Proof of Proposition 2.
1. Let us first prove that an optimal group subdivision satisfies the property described in the formulation of Proposition 2. Indeed, let us start with an optimal subdivision. Within each group, we can sort its \( g \) students in the increasing order of their grades; thus, every student gets assigned a rank in the corresponding group. Now, we can prove that for every two ranks \( r < r' \), a grade of a student of rank \( r \) is always less than or equal to the grade of a student of rank \( r' \) – even when they are from different groups. Similarly to the proof of Proposition 1, this can be proven by contradiction: if a grade \( d_j \) of a student of rank \( r \) is larger than the grade \( d_j \) of a student of rank \( r' \), then we can swap these two students and
improve the overall gain.

2. To complete the proof, we must show that for all subdivisions that satisfy the condition from the formulation of Proposition 2, the gain is the same. Indeed, the overall gain is equal to the sum of gains obtained in each group. Let us therefore calculate the gain in each group.

For $g = 4$, once we have $d_i \leq d_j \leq d_k \leq d_l$, the gain is equal to $\alpha \cdot (d_i - d_j) + \beta \cdot (d_j - d_i) + \gamma \cdot (d_k - d_i) + \delta \cdot (d_i - d_l) = 3d_i + d_k - d_j - 3d_i$.

Thus, the overall gain is equal to $3 \sum_{i \in L_3} d_i + \sum_{i \in L_2} d_i - \sum_{i \in L_1} d_i - 3 \sum_{i \in L_0} d_i$.

For a general group size $g$, one can prove, by induction, that once we have $d_{i_1} \leq \cdots \leq d_{i_g}$, then the gain of this group is equal to

$$(g - 1) \cdot d_{i_g} + (g - 1) \cdot d_{i_{g-1}} + \cdots + (2k - g - 1) \cdot d_{i_k} + \cdots - (g - 3) \cdot d_{i_2} - (g - 1) \cdot d_{i_1}.$$ 

Thus, the overall gain is equal to

$$(g - 1) \sum_{i \in L_{g-1}} d_i + (g - 3) \sum_{i \in L_{g-2}} d_i + \cdots + (2k - g + 1) \sum_{i \in L_k} d_i - (g - 3) \sum_{i \in L_1} d_i - (g - 1) \sum_{i \in L_0} d_i.$$ 

So, the sum does not depend on the subdivision – as long as the subdivision satisfies the condition from the formulation of Proposition 2.

The statement is proven, and so is the proposition.

**Proof of Propositions 3 and 4.** We have already mentioned, in the proof of Proposition 1, that optimizing the average grade is equivalent to optimizing the overall gain.

In model described by Definition 4, the gain coming from interaction between the $i$-th and the $j$-th students with $d_i < d_j$ is equal to $\alpha \cdot (d_i - d_j) + \beta \cdot (d_j - d_i) = \alpha' \cdot (d_j - d_i)$, where $\alpha' \equiv \alpha + \beta$. Thus, in the new model, the overall gain is described by the same formula as in the old model, but with a new coefficient $\alpha'$ instead of the original coefficient $\alpha$. Since the formula for the overall gain is the same, the optimal subdivisions are also the same. The propositions are thus proven.