



2010

Parts Of The Whole: Thinking about Variance: Standards, Targets, Tracking, and Other Thoughts

Dorothy Wallace

Dartmouth College, dorothy.wallace@dartmouth.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/numeracy>

 Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Wallace, Dorothy (2010) "Parts Of The Whole: Thinking about Variance: Standards, Targets, Tracking, and Other Thoughts,"

Numeracy: Vol. 3 : Iss. 2 , Article 9.

DOI: <http://dx.doi.org/10.5038/1936-4660.3.2.9>

Available at: <http://scholarcommons.usf.edu/numeracy/vol3/iss2/art9>

Parts Of The Whole: Thinking about Variance: Standards, Targets, Tracking, and Other Thoughts

Abstract

Variation is a natural result of any process, including education. Understanding how variation propagates and increases is necessary for designing educational interventions that work for the intended population. We show how common strategies such as setting standards and tracking can accidentally produce unintended and undesirable results due to the way variation moves through a system.

Keywords

education policy, quantitative literacy, numeracy, assessment

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Dorothy Wallace is a professor of mathematics at Dartmouth. She was 2000 New Hampshire CASE Professor of the Year, and the lead PI of the seminal NSF project, Mathematics Across the Curriculum. She recently finished a text in mathematical biology for first year students, *Situated Complexity*. She was a charter board member of the National Numeracy Network and is now co-editor of this journal.

Parts Of The Whole

A Column by D. Wallace

The problem of how best to improve the numeracy of a society is a thorny one, embracing the learning process of a single student but rising in scale to include the management and alteration of an entire system of education. With the issue of quantitative literacy always in mind, this column considers various aspects of the systemic workings of education, the forces acting on classrooms, teachers and students, and mechanisms of both stasis and change.

Thinking about Variance: Standards, Targets, Tracking, and Other Thoughts

In a previous column I talked about the idea of “statistical control” in a system, and some of the difficulties that arise if a system is not in control. In this essay I want to look at some of the issues that arise even when the system is in statistical control and is therefore relatively well behaved. The discussion here is general, but readers can keep in mind that the system we care about is education, the process is learning or whatever else happens in school, and the data points can be instantiated as students’ level of quantitative literacy. I will frame conclusions, when possible, in the context of numeracy.

The word, “variance”, derived from the same etymological roots as “vary” and “variation,” is a mathematical term. It is a number that is computed from a data set, describing the extent to which the data spread themselves around their mean. A large variance in a collection of data means that the numbers do not cluster around the mean too closely. A small variance means that they cluster tightly around the mean. If all the data points are exactly the same, the variance is zero.

If a process is in statistical control and we plot a lot of measurements of one of its sensors, we will see (by the very definition of statistical control) a picture like Figure 1. The mean is marked and also some cutoff lines that are computed from the variance. A system in statistical control will always place a predictable proportion of its data within these lines.

Variation is natural to any system. We must expect it, but it is better if we can predict it, and best if we can exert some control over it. A system as simple as the firing of a gun produces variation of output. Even the process of measuring the exact same quantity over and over produces variation in the measurements. The average measurement, the mean, is a pretty good indicator of how the system is doing, although incomplete. A well-behaved system in statistical control requires two numbers to describe its outcome: mean and variance. To use again

the firing of a gun as illustration, the mean describes where the gun is aiming, and the variance describes the error in hitting the mark.

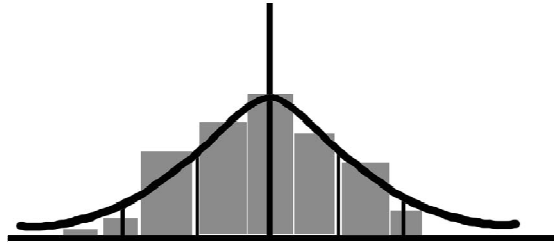


Figure 1. A system in statistical control.

What these two numbers signify to us depends on our point of view. For the sake of understanding, we must draw a sharp distinction between our all-too-human tendency to focus on individual data points and the pressing need to look at the system's output in aggregate. For example, any examination is some kind of measure of how the system is doing. A student who obtains a score well above average on an exam feels triumphant. A student who obtains a score below average may feel disappointed. Yet, if the system producing test scores is in statistical control, approximately half the students will score above average and about half will score below average. In fact, if you just test the same student over and over on a similar test, those scores will also exhibit variation from their mean: about half above, about half below. It is mathematically futile to attempt to make more than about half the population score above the mean.

Improving mean performance

The way we hope to improve education, then, is by moving the mean itself. A hundred years ago students entering college were expected to understand geometry. Now they are expected to understand algebra *and* geometry. Many students have taken calculus before entering college. The mean has shifted. There are two ways this might happen during a systemic change. These are pictured in Figures 2 and 3.

Notice that, in both cases, the mean has shifted upward. Yet, in the first case, the variance has remained the same while, in the second case, variation has increased dramatically. In all cases the system exhibits statistical control. There are many reasons why the scenario pictured in Figure 3 is more likely than the one in Figure 2. But there is one basic reason we would expect to see greater spread, and that reason is mathematical. Learning is a process and all processes have inherent variation. Asking for an improved mean usually implies more

learning, which is more of a process, leading to a greater variance in the resulting distribution, resulting in the famous “widening gap” between the top and bottom achievers.

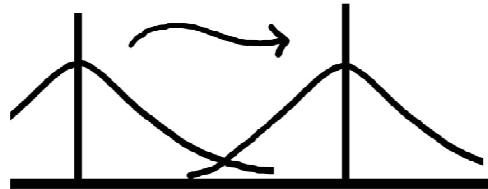


Figure 2. A shift in mean performance with no increase in variance.

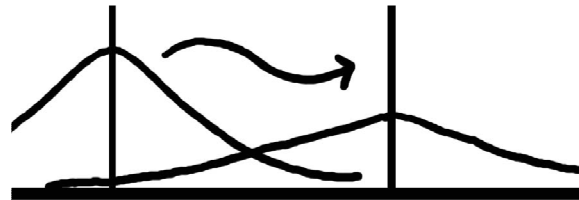


Figure 3. A shift in mean performance with increasing variance.

It is even possible to improve the mean while flattening the curve to the point where there are more data points in a region of poor performance than before. If it is hard to imagine how such a thing might come to pass, let me offer a hypothetical example. Suppose a class of students is working on reading. At any grade level, in any class, there will be variation among these students both at the beginning and at the end of the school year. Perhaps someone in the administration wishes to improve the mean score on a reading test for these students at the end of the year. As you can see, this example is completely hypothetical, as no one can provide two identical classes of students on which to test possible interventions. But perhaps the administrator has years of data to rely upon and a firm idea of what the mean and variance would be if no intervention were offered. A tutor is provided to the teacher in the hopes of enhancing overall performance of the students. The tutor, often a parent or other less experienced educator than the teacher, is given a handful of the poorest students in the class. The tutor does his or her best with these students, but in the end their performance

is pretty much the same as it would have been in the ordinary class. In fact, the teaching may even consist of repeated or reviewed lessons and pedagogy that did not work on the first try. Meanwhile, the teacher is free to push the remaining students a little bit harder and set sites a bit higher, knowing none will suffer for it and many are likely to benefit. What will be the final outcome of this intervention? The poorest students, although not harmed, are where they would have been otherwise. The better students have improved. The mean for the whole class has gone up but just as many students score in the low part of the original distribution as before. The variance has grown.

Taken in isolation, such an outcome may appear to be a complete triumph. But, the next year, the teacher who gets these students is faced with a much wider variation in student background than otherwise. At some point, if the variation gets too big, the teacher cannot teach effectively. Any strategy devised is likely to work only with some fraction of the students. What started out as an inexpensive, helpful intervention may result ultimately in one that is necessary merely to have a functioning classroom. In fact, the more one thinks about classroom strategies for improving test scores, the harder it becomes to imagine an intervention that increases the mean but does not increase the variance. It is equally clear that the ever-increasing variation among students as they pass through the educational system is also natural, and cannot be solved by any single teacher acting in isolation. It is a systemic issue.

Accidentally lowering the mean

Of course, the mean performance of a student population can also be adjusted downwards. It is possible to achieve this as an accidental by-product of a well-meaning intervention. So it is appropriate in this essay to discuss the way in which establishing standards of educational achievement can accidentally lower student performance. In order to understand how such a perverse thing could happen, it is necessary to remember the gun of the previous essay, and what it means to aim it. The output of a gun in statistical control is a set of data points clustered predictably about a mean. The gun is “aimed at” the mean. To change where the gun is aimed means changing the mean of the data set.

What, then, are standards? Standards are set, not in response to the mean of a population, but in response to the lower end of the distribution. A standard is supposed to be something nearly everybody can achieve. Almost half the population will never achieve the mean score on any test, so the standard is set to determine where the bottom of the distribution ought to be. In and of itself, this is an admirable attempt to increase the mean by the very method of reducing variation, bringing the bottom of the distribution up a little bit higher.

The potential problem is interpretation. The principal and the teachers have been given a lower bound on what students should be able to do, but they haven’t been told what the average student ought to be able to do. It is very easy to

misconstrue the “standards” as the “target”. In fact, in the absence of a stated target it is hard to do otherwise. If the educational system starts aiming its outcomes at written standards that were intended as a lower cutoff, the mean performance of students is bound to decrease. Figure 4 attempts to illustrate this point.

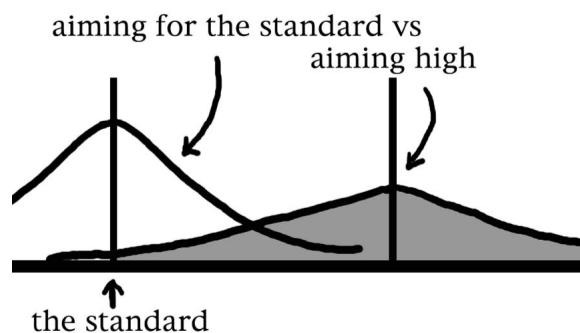


Figure 4. Improving the mean at the cost of increased variance.

An example of attempting to offer a target is the goal of “algebra for every student.” This goal is clearly a target for which to aim. Based on our discussion, what would be the likely outcome of taking this goal seriously? Probably nearly everyone will see some algebra. There will be some level of algebraic skills and understanding that will represent the average of the population. Some will do far better than the average and some will do far worse. There is no way to predict how big the variance would be. Nor is there any indication, because of the absence of standards, what kind of baseline understanding would be acceptable. In the absence of a strategy for controlling variance, such a goal is likely to result in an improvement of some mean measure of algebra understanding, while simultaneously making the system harder to steer because of increased variance within it. As part of a systemic strategy including these considerations, it could be a powerful force for improvement. Without consideration of the variance, it could well result in worse mathematics education for some portion of the population.

How variation propagates in a system

If we consider education as an aggregate of smaller processes happening in every classroom and at every kitchen table, we must realize that each of these smaller processes has its own natural variation in output. And so we come to one of the main mathematical properties of variance. When many independent processes join together to form a larger system, their individual variances sum. The

variance of the system as a whole will tend to be the sum of the variances of all the factors contributing to it.

The process of teaching fourth grade reading depends upon the various inputs to the fourth grade experience, but it does not depend very much on the details of the process of teaching third grade reading. In other words, the actual processes of third and fourth grade are, for the most part, fairly independent of one another. When independent processes are combined to form a larger process, the variances of the two processes sum to give the total variance for the combined effect. In other words, unless extreme steps are taken to prevent it from happening, the variance in quality of student performance will increase every single year. This observation is mathematical, and does not depend on any particular educational approach or pedagogical technique, although every choice of approach or technique has some kind of effect on the variance as well as the mean. If we pick a moment, such as the end of high school, to measure the output of our educational system, the variation we observe will be, more or less, the sum of all the variations of all the small processes that contribute to the whole educational experience.

One can see this effect, for example, in mathematics education. In kindergarten there is a very small variation in ability and knowledge of incoming children. Some can count well, perhaps a few can add small numbers. Some do not yet count well. This is a very small difference in ability. By fifth or sixth grade everyone will be able to count and add small numbers, yet the variation in skill level will have increased to the point where some children master addition of fractions fairly quickly while others remain confused for the entire year. By high school, it is no longer possible to keep everyone together. Some will have advanced beyond basic algebra while others are still unable to figure percents and fractions, add decimals, or estimate. At the entrance to college we see wide disparity, from vast offerings of remedial courses such as “college algebra” to students who have passed out of one or more semesters of calculus. At this point not even a single institution of higher learning would attempt to address the full spectrum of student background. Instead we have two-year colleges, four-year colleges, universities, elite private institutions, and so forth. Not even an entire institution can deal with the variation in the mathematical ability of eighteen-year old adults.

In view of the preceding discussion, we should not be upset about such a state. It is not the fault of any particular part of the system. In fact, it is at least in part an inevitable result of our continual attempts to raise the mean level of student performance. Our best students are doing better than they were fifty years ago. As we have seen, it is relatively easy to devise methods for raising the mean performance at the expense of increased variation. Evidently, this is what has happened over the last fifty years, because while the mean performance has gone up, the variation among entering students has grown even more. Many of the readers of this essay may fondly remember a time, not more than thirty years ago,

when the decision to go to a two-year college before university was often a purely economic one. One would expect to obtain a similar education to the first two years of university, after which one would transfer and expect to be on a par with other juniors and seniors. Subsequent increased variation in the “preparation” of incoming students has forced the junior college down a different path from the university, and the two experiences are not really comparable any more.

Strategies for dealing with increasing variation

The natural human response to variation of this sort is to attempt to compensate for it in some way as students move from class to class and year to year. Special education, constant adjustment of standards, charter schools, tracking—all of these are attempts to deal with increasing variation among students. Some of the new private schools feature a “core curriculum” that, by feeding all students exactly the same material and requiring exactly the same mastery of it before graduation, would ostensibly keep variation down somewhat. Special education would increase the performance of disadvantaged students, thus bringing up the bottom of the distribution and reducing the variance, a goal that is achieved when the student can be “mainstreamed” into a regular classroom. Tracking does not seek to reduce variance, merely to cope with it by separating the students into multiple strands, each with smaller variation in it.

There are quite a few pitfalls to avoid with these kinds of interventions. The best intentions are not enough to guarantee that a change of strategy will have the desired effect. Deming offers a delightful example of this in the form of an experiment that is an oversimplification of human behavior, to be sure, but still illustrative of the possible side effects of good intentions. In the famous “funnel experiment,” small metal beads are dropped through a funnel onto paper backed by carbon paper, so that the location where they land is recorded. If the funnel is stationary, the marks will exhibit all the properties of a process in statistical control, clustering around the mean in the usual way. One can take measurements and compute the variance of the population. If one is unsatisfied, one can attempt to reduce the variance by dropping the beads through the funnel one at a time, adjusting the funnel every time a bead is dropped. There are various recipes for adjusting the funnel, corresponding to different kinds of responses that people might make to a situation. One of the most reasonable is to compensate for the “error,” the distance of the last bead from the proposed mean, by moving the funnel in the opposite direction of the error for a distance equal to the error. In other words, if the funnel is pointed at a particular location and the bead falls one centimeter to the right of where it is supposed to, then move the funnel one centimeter to the left. Measure and compensate after every drop. Because this is a specified algorithm for compensating for variation, it can be analyzed mathematically as well as run experimentally. Either way, the result is the same. By trying to compensate for variation in this way, the system is changed so that

the variance actually doubles as a result, although the mean stays the same. Unwittingly, and with the best intentions, the “problem” is made twice as bad.

Deming uses this experiment to make several points. First and most importantly, common variation, as exhibited by a system in statistical control, will not be reduced by compensating after each consecutive measurement. The result will always be an increase in variation. An example of this kind of behavior in education would be adjusting resources available to the teacher each year depending on the performance of the previous year’s class. Choice of difficulty in textbook from year to year based on the previous year’s class would also be an example of such an algorithm. The only way to innovate without running the risk of increasing variation is to base decisions on aggregate data compiled over many years running. If such data are not available, the next best thing is to base decisions on a sound theory that takes both the mean and variance into account.

Deming’s second observation about the funnel experiment is that it shows the inevitable effect of individuals attempting to reduce variation stemming from a common cause, while working completely within the system. These individuals are said by Deming to be “tampering” with the process. In doing so, they introduce a factor that actually increases the variance. To reduce the variance, Deming claims, one must work outside of the system to change the system itself. Deming would place such a goal completely outside the capabilities of individuals in the system such as teachers and students, placing it instead in the hands of those whose job is to manage the system as a whole.

Yet, viewed from the perspective of the individual teacher, the entire point is to move the average knowledge of students from one place to another during the course of the year. Clearly this is not the same as improving mean performance at the end of the process from year to year, a problem best left to larger forces, as in the example of tracking. But the teacher faces the same statistical issues as the management, on a smaller scale.

Students come into the class in the fall with varying background and natural ability in any given academic area. There is some average knowledge base with which the teacher must work. The teacher will help all students learn new things during the course of the year. Any process by which students learn will have an intrinsic variance of its own. If the original variation in the student population is not too large, the teacher can effectively teach new material with the inevitable result that the variation in performance at the end of the year will be slightly larger than at the start. If the variance in the entering class is too big, the teacher may not be able to teach everybody equally well. Even if the teacher is experienced and dedicated to reaching every single student, the variance at the end of the year may be so large that some are not ready for the next grade. It is not necessarily the fault of the teacher or students (although there are surely better and worse teachers and students), it is merely the way variation moves through a system.

It is easy to see why teachers desire and, in fact, need to teach a population of students whose variation in background is kept within limits. We may desire, or at least be able to run a country with, an adult population as widely varying as the flowers of the field. But a single classroom is a habitat that can respond effectively to the needs of those in it only as long as they are reasonably similar in needs. Society and the educational system it supports have long accommodated to this necessity.

Historically, the most common means of enforcing a constraint in variance was simply to eject less promising individuals from the educational process at a certain age. Removing the bottom of the distribution always reduces variance while increasing the mean. The business of education was easier then, but at a certain cost to society that we are unwilling to pay today. One of the author's elderly relatives describes her education by stating that she "went to school to eat lunch." This person, who completed her education at eighth grade (without algebra, we might note), is perfectly bright. Today we would not find her education acceptable and we would send her all the way through college, still eating her lunch. Because learning algebra requires more perseverance than eating lunch, this person would be a prime candidate for remedial courses at the two-year college. She would certainly contribute to the difficulties of variance described here.

Aiming for calculus

Another example of removing the bottom of the distribution is by establishing stiff prerequisites. Science departments at the college level routinely require a certain level of mathematics before entering the major. During the Sputnik years, when everyone wanted to be a scientist, the mathematics classes were explicitly depended upon to cull the population of potential scientists down to the very top mathematics students. The result was a consistently high mean and relatively low variance of knowledge among entering science majors. A tightly orchestrated calculus sequence, where prerequisites were delineated carefully and strictly enforced, proved to be an excellent solution to this problem.

Now that there is a perceived shortage of science, mathematics, and engineering majors, the nation would like to have more people going into those fields, but at no cost to the quality of their education. In other words, we want a bigger population of scientists, but without reducing the mean performance of the population or increasing the variation. For the last fifteen years people have somehow expected this to happen without any change in the standard course sequence or prerequisite structure. All improvements were supposed to come from alterations within existing courses. The Calculus Reform Initiative sponsored by the National Science Foundation was a prime example of this expectation. Without changing the preparation of incoming students, the curriculum, the course sequence, or the structure of the science major into which

these students were supposed to go, the pedagogical events inside one or two courses were supposed to alter a societal trend. Individual teachers were expected to create a large-scale systemic improvement. Although within the context of individual courses some improvements could clearly be claimed, the effect on the system as a whole has been negligible, exactly as Deming would have predicted.

A side effect of the push to create a large population of students who have taken calculus is the concern raised by members of the business community and others, who point out that calculus is not really needed (yet) in the business world. Because the majority of people who take calculus do not go on to become scientists, but instead function more or less in the business world, many now question whether the time spent learning calculus might not be more profitably spent on other types of mathematics. For this, among other reasons, the “quantitative literacy” movement is taking hold in higher education as well as elementary school. Its promoters claim that time would be better spent reinforcing earlier concepts commonly used in business, such as estimation skills, data analysis, and the algebraic skills necessary to design a spreadsheet. While there is every reason to believe that students will need to improve these skills, making them the explicit aim of college mathematics is a dangerous thing, precisely because they represent a standard that most college students ought to achieve. Requiring all students to master these skills can do no harm. On the other hand, the suggestion to replace calculus with lower level mathematics that all students ought to learn is a perfect example of setting a standard that is lower than the mean performance of the student body, and then aiming for it. The result is sure to be a reduction in overall performance.

Aiming for the “standard” at each level of education will necessarily result in it becoming increasingly harder to meet succeeding standards as the level goes up. As people move through the system the variation in their performance will increase, with the lowest achievers falling further and further below the “standard” at which their education is aimed. Unless a target is explicitly given for a particular educational endeavor, the standards that have been set will be used as the target for practical purposes. It is critical that the very different functions of “standard” and “target” be recognized and that educators and administrators do not confuse the two. “All students should have reasonable estimation skills” is an example of a standard, whereas “Algebra for all” is an example of a target.

Tracking

Another common response to high variation among students is tracking. Tracking is a managerial strategy whereby a population of students with widely varying preparation is divided into two or more groups according to some measurement of their level at the beginning of the course, course sequence, or academic year. Suppose, for the sake of illustration, that we pick a cutoff point and divide the population in half. The higher achievers now have a higher mean and reduced

variance in their population. Their teacher can reasonably attempt to get them a little farther than they otherwise would have if the population remained large. The lower achievers have a lower mean than the original complete population, but also a reduced variance. A good teacher can do a better job with them as a result, perhaps improving their mean score at the end of the year over what it would have been had the class been kept together.

So far this looks good. The very top students will clearly benefit. The very bottom students will probably also benefit, because assumptions about the background of the group they are in are now somewhat closer to their reality. Students near the cutoff may not fare so well. A student tracked into the higher level group will probably learn at least as much as if the class had been kept together. But a student near the cutoff who is tracked into the lower level group will almost surely learn less than if the class had been kept as a whole. Because there is always a certain error in any measurement, a student who scores near the cutoff on a diagnostic test could equally well score above or below it, depending on the wording of a single problem, or personal circumstances on the day of the test. If the cutoff chosen is the mean score on the diagnostic test then it is likely that a large number of students fall into this category. Placing these students in one track or the other is, then, fairly arbitrary.

This bit of unfairness resides on the level of the individual. There are also systemic issues with tracking that need to be scrutinized. When the population is divided in half, say, to form two tracks, each of the two smaller populations becomes an educational unit. The teacher, indeed the whole system, is going to aim the two new units very differently. We started out with one population with certain preparation. If the system were in statistical control (the best situation for managing a system) then the distribution of necessary background knowledge would look something like Figure 5.

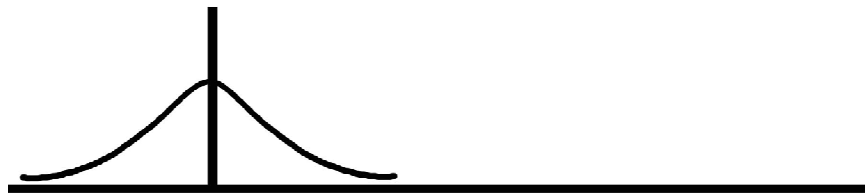


Figure 5. A system in control (again).

If the population could have been kept as a whole and taught effectively, the result would have looked like Figure 6.

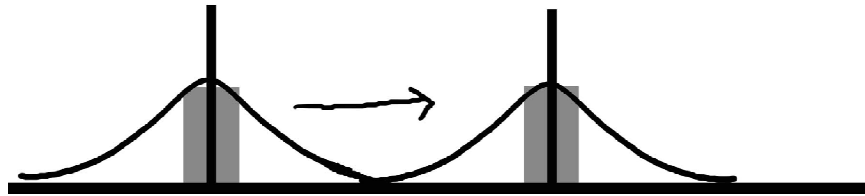


Figure 6. A whole population aimed at a single goal shifts the distribution upward.

The distribution on the right would have slightly larger variance, but look basically the same. If each of the two sub-populations is now treated as a unit and aimed for different, although reasonable, levels, what will happen? After a while, if each of the two processes is in statistical control (the best situation!) the achievement levels of the two populations will eventually look like Figure 7.

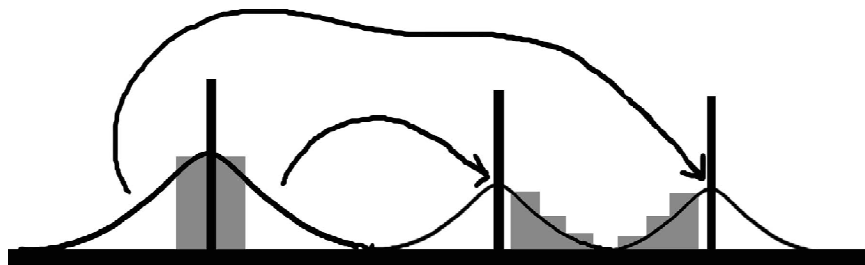


Figure 7. The population is divided and aimed for different goals.

Students near the cutoff, represented by the shaded gray boxes on the left distribution, began the year looking fairly similar to one another. These students are now very different in achievement, separating from each other like the gray boxes under the two distributions on the right. This difference is not a product of bad teaching, or of the abilities of the students. It is purely a result of the system. The picture makes it clear why it is so difficult for students to change tracks after a while. The highest achievers in the lower track continue to move farther and farther from the mean performance of students in the upper track. The lowest achievers in the upper track will gain nothing except good grades by moving into the lower one. Taken as a whole, the system that contains both populations of students is no longer in statistical control. There is nobody near the mean performance of the total population, although each of the two smaller populations may well be in statistical control. The matter is critical if the subject in question affects college entrance decisions, such as writing, reading comprehension, or

math. These two populations of students, if kept apart long enough over the course of many years, will diverge to the point where they actually require different kinds of colleges by the time they graduate high school because no sensible set of entrance requirements could be given for both populations simultaneously.

It is important to point out in this example that often tracking is instituted precisely because the variation among students has grown so great that it is impossible to teach them as a single population, even though their preparation reflects a process in statistical control. It should be clear from the discussion above, though, that there are ever-present side effects of creating tracks that should be minimized if at all possible. The random effect on students near the cutoff points could be reduced, by placing the cutoffs fairly far from the mean. A single track consisting of a wide range of students around the mean performance will cause less random havoc than placing a track where the cutoff score falls near the mean, as illustrated in Figure 8.

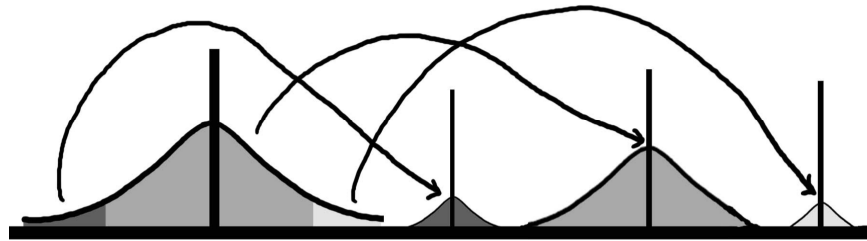


Figure 8. Tracking with less room for error.

Tracking that is instituted late in the educational process will have less irreversible results than tracking that is instituted early. A difference in preparation of only one course or two at the beginning of a student's college career will have far less impact on that student than a similar difference in preparation at the start of high school. It takes time for the different tracks to diverge from each other, and the longer the tracking continues, the greater the difference will be.

Finally, the observation that the tracks move farther and farther apart is not a necessary feature of the system; it is merely how it usually works. If the teachers of students in the lower tracks could be convinced and supported by the system to aim for the exact same level of achievement as the higher track, then the tracks could, in theory, stay close. Such a stance on the part of teachers and administrators is counterintuitive, yet the expectations the teacher places on the

students are a critical factor in their performance. It is even possible that variation could be reduced by this method, because each teacher is working with a more homogeneous population of students. It would be interesting to know of any examples where tracking was observed to produce such an outcome.

Controlling variation in a process

It might be useful at this point to summarize the general advice that Deming himself gives about controlling variation in a process. To implement his advice it is necessary to have data about each step in the process, but we will assume such data might become available. First of all, at each step the process needs to be in statistical control. Suppose that, at some grade level, scores on diagnostic tests are shown to become erratic or bimodal, or other evidence is given that the process has jumped out of control. Deming would say that this is a result of a special cause. That cause should be looked for and found. Until it is, no subsequent part of the process can be analyzed effectively. If the whole process is in control, the points in the process where variation increases most dramatically ought to be analyzed. The source of variation in these cases is common, and resides typically in the structure of the system as a whole. These are the places where attention ought to be given first. Finally, because early variation propagates throughout the entire system, special attention ought to be given to the early processes, in this case, the early grades. None of these suggestions tell us how to improve mean performance, however, because that is a problem particular to the process itself and not visible from its statistics.

The examples given here are intended to illustrate the value of studying any existing or proposed system intervention in light of its effect on both the mean and variation of outcome within the student population. In fact, as the example of tracking shows, one can sometimes get a useful theoretical picture of how the shape of the entire distribution of results might change because of an intervention. These graphs, arguments, and general principles are tools that those who manage the entire educational endeavor can use to make better choices on behalf of all students and fair assessments of those who teach them. Those of us promoting quantitative literacy as a goal of education need to be aware of how variation responds to basic interventions such as setting standards and targets, and the assessment of student and teacher performance.