



A Numerical Indicator of Student Cognitive Engagement and Mathematical Growth

Mercedes A. McGowen ^{1*} , Gary E. Davis ² 

¹William Rainey Harper College, Palatine, Illinois, USA

²University of Massachusetts Dartmouth, Dartmouth, Massachusetts, USA

*Corresponding Author: mercmcgowen@sbcglobal.net

Citation: McGowen, M. A., & Davis, G. E. (2022). A Numerical Indicator of Student Cognitive Engagement and Mathematical Growth. *International Electronic Journal of Mathematics Education*, 17(1), em0669. <https://doi.org/10.29333/iejme/11473>

ARTICLE INFO

Received: 22 Aug. 2021

Accepted: 9 Nov. 2021

ABSTRACT

We discuss and examine a numerical indicator—the individual gain—of students' engagement and mathematical growth in relation to an instructor's course aims and goals. The individual gain statistic assesses the fractional amount an individual student improves initial-test to final-test. We argue that an initial-test score and a final-test score, if the two tests are related to each other and to a course focus, can provide a numerical indication of a student's engagement with the goals and aims of the course and the extent to which a student was prepared to work toward those goals. Results on the distribution of individual gain for students in two-year college developmental mathematics courses and in sections of a course for pre-service elementary teachers are discussed. We detail and discuss advantages of the full distribution of individual gain, particularly for allowing statistical inference for differences, compared with Richard Hake's use of mean gain of reform classes in undergraduate physics. Other instructional benefits of using the gain statistic to examine distribution of individual student gains include: a pre-test formative assessment at beginning of instruction, providing an instructor with data for specific, targeted remediation; and planning information that informs an instructor for the effectiveness of instruction for students in that cohort.

Keywords: assessment, pre-test, post-test, gain, engagement, growth

INTRODUCTION

We examine the issue of assessing student growth in mathematical thinking in relation to an instructor's course aims and goals for pre-service elementary teachers and college developmental algebra students, using the usual assessment instruments and student data available to an instructor. What does a grade or average tell us about what, specifically, an individual student in a college mathematics class has learned? Does the assessment inform an instructor about each student's mathematical growth and the extent of engagement with the instructor's explicit articulation of course aims and content? Consider the following cases:

1. Two students in an introductory algebra course had quite different scores on a course pre-test: Student A had a pre-course score of 59% while student B's score was 82%. Over the course of the semester, student A's test scores were, in order: 68%, 76%, 91%, and 95%, for an average of 83%. Student B's test scores were, in order: 78%, 81%, 87%, and 84%, also for an average of 83%. What, if anything, do the averages of 83% tell us about the engagement and mathematical growth of these two students over the semester?
2. Comparison of final common examination scores and course grades of 321 college students enrolled in 20 sections of a developmental introductory algebra course over one semester were examined for consistency between course grades and departmental final examination scores, revealing an unexpected and startling degree of overlap in course grades for a given final examination score. Several students who received a grade of 83% on the common departmental comprehensive final examination were given course grades of A, B, C, and D. Students who received a grade of 80% on the final examination were given course grades of A, B, C, or F. A score of 72% on the departmental final examination also had a course grade of A, B, C, D, or F, as did final examination scores of 63% and 57%.
3. An analysis of pre-test and post-test data of students in an Intermediate Algebra developmental course revealed that the significant change in the overall pre and post-test class responses was due primarily to differences in the pre-test and post-test responses of the top four students in the class.

The complex nature of the grading process and the meaning and interpretation of grades has been critically examined by, among others, Bouyssou et al. (2006), Brookhart (1999), Burkholder (2015), Close (2009), Foegen (2000), Foegen et al. (2007), Hammons and Barnsley (1992), McCrickerd (2012), Milton et al. (1986), Moskal (2000, 2002), and Willingham et al. (2002). Practices

that can enhance learning are detailed in reviews of research on assessment practices (Fuchs & Fuchs, 1986; Crooks, 1988; Wiliam, 2011; Black & Wiliam, 2009, 2018).

Components contributing to student grades in mathematics classes include examinations, quizzes, homework, projects, research summaries, and in-class presentations. Instructors have varying reasons for assessment. Some of these include: as a summative assessment of learning and retention of knowledge, a check on student application to assigned homework, a component of a multi-faceted assessment suite, or a normative check on teaching and learning standards across several sections of a course. As we pondered discrepancies between grades and assessment we were led to consider how individual students might demonstrate mathematical growth over the course of a semester, and how that growth might relate to an instructor's explicit articulation of course aims and content. For us, students of mathematics demonstrate growth in mathematical thinking when they move from a procedural, mainly instrumental, approach to mathematical problems, in which their problem solving attempts are limited by conditioned application of rote recalled formulas and procedures, to a position where they are, at least, beginning to demonstrate some or all of the following:

- Flexible mathematical thinking. In particular, answering paired questions to indicate ability to reverse train of thought, or seeing the relationships between a series of related questions on a given topic.
- Critical thinking as they analyze problems, shown via written work and justifications.
- Reflection on what they know or how they know it, via written reflections and self-assessments.
- Making conjectures and presenting arguments to support or explain conjectures.
- Synthesizing key mathematical ideas and problem-solving approaches by applying them to diverse problems and by exploring interconnections using appropriate technologies.
- Communicating mathematics accurately, verbally, in writing, and in the use of various manipulatives and representations.
- Using a variety of tools, physical models, and appropriate technology to demonstrate an understanding of concepts, relationships, and applications.

Assessing Mathematical Growth: Individual Gain and Hake's Mean Gain

We argue that a statistic derived from initial-test and final-test scores—the individual gain, defined below—provides us with a measure of the extent to which an individual student engages with mathematical content and an instructor's explicit goals and aims related to deeper understanding. Student written statements, as indicative of the degree of engagement with the mathematics content and a changed attitude towards mathematics, are examined in relation to the individual gain.

Prior to our formulation of the individual gain, and unknown to us at the time, Hake (1998) introduced the mean gain—denoted $\langle \text{gain} \rangle$ —for classes, consisting of over 6,000 undergraduate physics students in total, who were given a pre-test and a post-test in undergraduate physics. Hake (1998) calculated a mean, denoted $\langle \text{gain} \rangle$, from the mean scores on a pre-test and on a post-test:

$$\langle \text{gain} \rangle = (\text{mean post-test \%} - \text{mean pre-test \%}) / (100\% - \text{mean pre-test \%})$$

We defined the individual *gain*, initial-test to final-test, as follows (Davis & McGowen, 2001; McGowen & Davis, 2001, 2002):

$$\text{gain} = (\text{final-test\%} - \text{initial-test\%}) / (100\% - \text{initial-test\%})$$

Note that *gain* is undefined if the initial-test score is 100%—a situation we have not encountered in the data here, or in similar data collected over many years.

There is, theoretically, a difference between Hake's mean $\langle \text{gain} \rangle$ for a cohort of students and the mean of individual student gains over a cohort. In practice we have found, over many years, that these two average quantities differ only a little. Bao (2006) has discussed this theoretical difference via simulation. The mean of the individual gains is one of the many useful summaries that can be obtained from the full distribution of individual gains and approximates Hake's mean $\langle \text{gain} \rangle$. The full distribution of individual gains allows us to use standard methods of statistical inference to make statistical comparisons across different classes, instructors, and methods and focus of teaching. Further, a smooth kernel distribution approximation to the distribution of individual initial and final test scores can often make Hake's mean gain appear more visually interpretable. For example, as shown in **Figure 1**, for a class of 30 intermediate algebra students at a 2-year college, we can visualize smooth approximations to the distributions of initial and final test scores, along with means of those distributions.

It is visually clear from this image that there has been a marked shift from the mean initial-test score of 0.21 to the mean final-test score of 0.61. Moreover, the full distributions, rather than the summary means, allows us to infer that the difference in means is statistically highly unlikely to have happened by chance: a *t*-test for the difference in means for initial-test and final-test scores returns a *p*-value approximately equal to the probability of obtaining 46 heads in a row from tosses of a fair coin. Further, we can calculate a 99% confidence interval for the initial-test mean: (0.17, 0.26) and the final-test mean (0.53, 0.70), as well as for the difference in means: (0.30, 0.49). This provides us with more information than simple point estimates for the means or the difference in means.

Both the mean $\langle \text{gain} \rangle$ and individual *gain* (also referred to as “normalized learning scores”) have been used widely in science education, particularly in relation to biology and physics teaching and learning (Bao, 2006; Chi et al., 2018; Cobern et al., 2010; Concetta Capizzo et al., 2006; Docktor & Mestre, 2014; Fadaei, 2019; Hake, 1998; Knight et al., 2013; Marx & Cummings, 2007; Redish et al., 1997; Setiawan & Kudus, 2020; Smith et al., 2011), but relatively rarely, to date, has it been used in mathematics education.

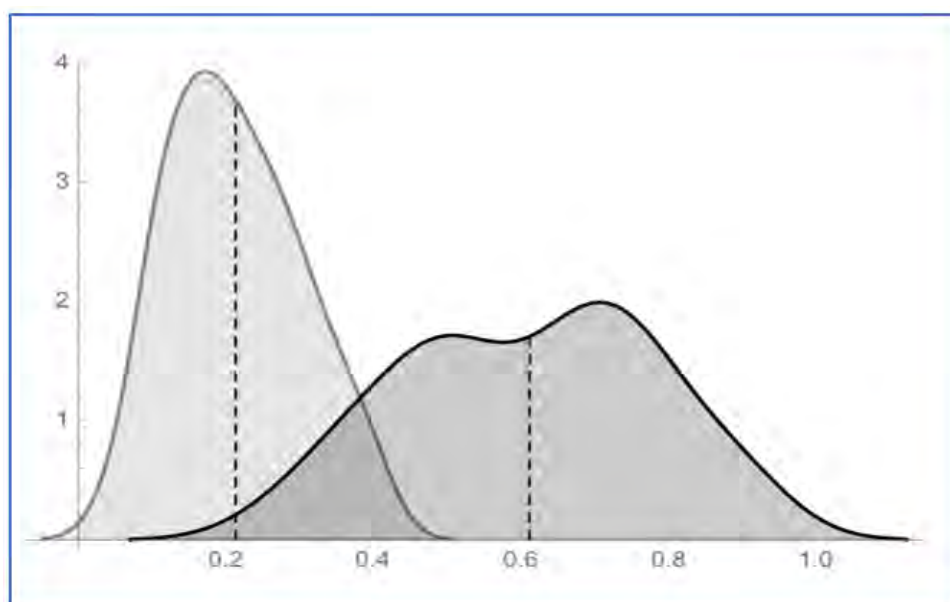


Figure 1. Smooth approximations to distributions of initial-test scores (left) and final-test scores (right) for a class of intermediate algebra students ($N = 30$), with mean scores shown dashed

The individual gain bears no necessary relation to the initial test score and generally has low correlation with it. We argue the individual gain is a good numerical indicator of the extent to which students have a willingness to change their ways of thinking mathematically, to become more flexible mathematical thinkers, and to engage fully with articulated goals of instruction. The individual gain statistic, calculated for each student from an initial-test score and a final examination score, can, if the two tests are related to each other and to the course focus, provide a numerical indication of a student's engagement with the goals and aims of the course and the extent to which a student was prepared to work toward those goals. We will argue the z-scores of individual gains provide a measure of student growth as well as an indicator of instruction effectiveness.

In **Appendix A**, we discuss the individual gain statistic in relation to other relative change functions (Törnqvist et al., 1985; Bonate, 2000; Dimitrov & Rumrill Jr., 2003).

Comparison of instructional methods and content

Hake (1998) studied the mean gain for classes consisting of over 6,000 undergraduate physics students in total, and concluded that, generally, high mean gains were associated with reform classes, whilst low mean gains were associated with more traditional lecture-style courses. Hake's mean gain is useful in assessing the instructor's impact on student learning and growth, while the individual gain is useful for assessing a student's learning and growth in relation to an instructor's explicit goals and objectives. The results of our analyses agree with Hake's conclusion that instructional methods and content that promote active student engagement have a positive impact on student learning and growth.

Descriptive Analysis and Statistical Inference in Mean Gains

The use of individual gain, rather than the summary class or cohort mean gain, allows descriptive analysis of the distribution of gains as well as statistical inferences of the difference in mean gains between cohorts. We use standard probability density function histograms, along with their smooth kernel approximations, to visualize and compare distributions of individual gains. Readers should note that smooth kernel approximations generally overrun discrete histograms, so for example smooth histograms will appear to indicate that there are examples of individual gains greater than 1, whereas this is an artifact of approximating the discrete distribution of gains.

We utilize standard t -tests for comparison of mean gains for different cohorts of students, and we report the p -value from the t -test, as well as the negative base 2 logarithm of the p -value, which is an indication of how many heads in a row, from tosses of a fair coin, we would expect to see with that probability if the null hypothesis of no difference in mean gains were true. In some cases, a t -test for comparison of mean gains was not appropriate due to significant deviations from normality of the distribution of individual gains. In those cases, we utilized a bootstrapping method to estimate the probability of a difference as large as we see if the null hypothesis of no mean difference were true. Cohen's d was utilized to estimate the effect size of a difference in mean gains between cohorts. Correlations between gain and initial and final test scores were examined using simple linear regression.

Individual gain and fractional increase

A common way for a teacher to assess a student's change, from one test to a later similar test, is to calculate a fractional, or proportional, increase in test scores (Bonate, 2000):

$$\text{fractional increase} = (\text{final-test score} - \text{initial-test score}) / \text{initial-test score}$$

As shown in **Figure 2**, the distribution of fractional increase scores for 155 pre-service elementary teachers is strikingly different from the distribution of individual gains. The significant difference in the scales on both axes can be seen in the figure.

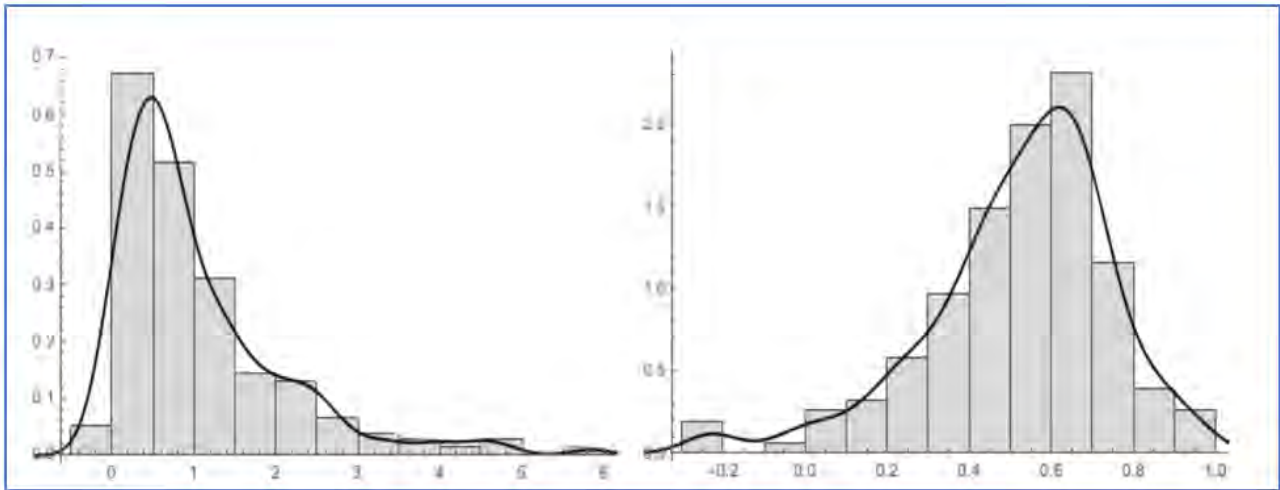


Figure 2. Histograms with smooth kernel approximations, of distribution of fractional increase (left) and individual gain (right) (N = 155).

The fractional increase and individual gain are not independent variables, but they are generally only weakly correlated. For the initial-test /final-examination data from a study, with 155 students, the correlation between fractional increase and individual gain was only $r^2 = 0.20$. The fractional change, (final-test score – initial-test score)/initial-test score, correlated quadratically with initial-test scores ($r^2 = 0.88$). In our context, therefore, the gain function provides significant extra statistical information beyond initial-test scores.

Correlation of individual gain with test scores: Initial-test scores

Generally, individual gain has low correlation with initial-test scores. **Figure 3** is a plot of individual gain versus initial-test scores for 155 pre-service teachers in the pre-service mathematics content course.

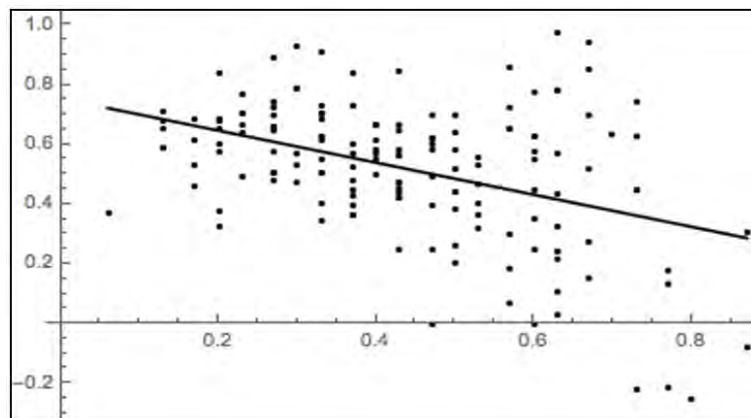


Figure 3. Individual gain versus initial-test scores (N = 155 & Pearson's $r^2 = 0.174$ for linear regression)

This is typical of plots of individual gain versus initial-test scores, and is in accord with Hake's observations on his mean <gain> (Hake, 1998).

To put this correlation in context, we compare it with the correlation between initial and final test scores for the same cohort of students as shown in **Figure 4**. Initial test scores are generally weak predictors of final test scores. In this case, a linear regression model has $r^2 = 0.164$ (N = 155), so only approximately one-sixth of the variation in final test scores is linearly modelled by variation in initial test scores: fully five-sixths is not.

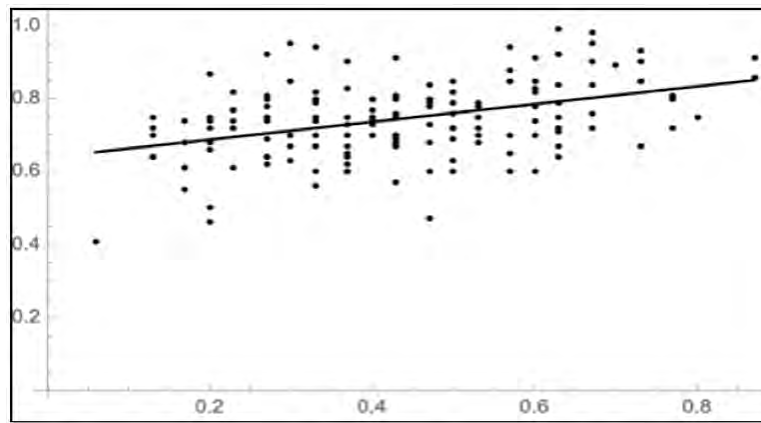


Figure 4. Correlation of final-test scores with initial test scores (N = 155 & Pearson’s $r^2 = 0.164$ for linear regression)

Individual gains and below average initial-test scores

The correlation between individual gain and initial test scores is of the same order of magnitude, in this instance: Fully 83% of the variation in individual gain is not linearly modelled with variation in initial-test scores. As depicted in **Table 1**, median initial-test for combined group (N = 155) was 0.43.

Table 1. Mean of individual gains for groups X and Y

Group	N	Gain	
		Mean	95% CI for mean
X	73	0.60	(0.57, 0.64)
Y	82	0.44	(0.39, 0.50)

A *t*-test gives $p < 0.000004$ so, under the null hypothesis that there is no statistical difference in the mean gain between group X and group Y, we would see the difference we do see with probability equivalent to tossing 18 heads in a row from a fair coin. We conclude that group X had a statistically highly significantly greater mean gain than did group Y.

We estimate the effect size in difference in mean gains using Cohen’s *d*, which we calculate to be 0.75. This borders on a “large” effect size, and 88% of group X gains are above the mean of group Y gains. This seems to add weight to the belief that lower initial-test scores generally entail higher individual gains. However, this is a statistical conclusion and *there are significant and meaningful exceptions to this general rule*, as we see below.

Final-test scores

Individual gain generally increases overall with final-test scores, with varying degrees of correlation. For example, for the same cohort of 155 students as above, a plot of individual gain versus final-test scores shows a moderate linear correlation as shown in **Figure 5**.

We have found, in practice, the linear correlation between individual gain and final-test score can vary widely, with r^2 ranging from a low of around 0.3 to almost 1. It is not clear what, if anything, a particular degree of correlation of gain with final-test score signifies.

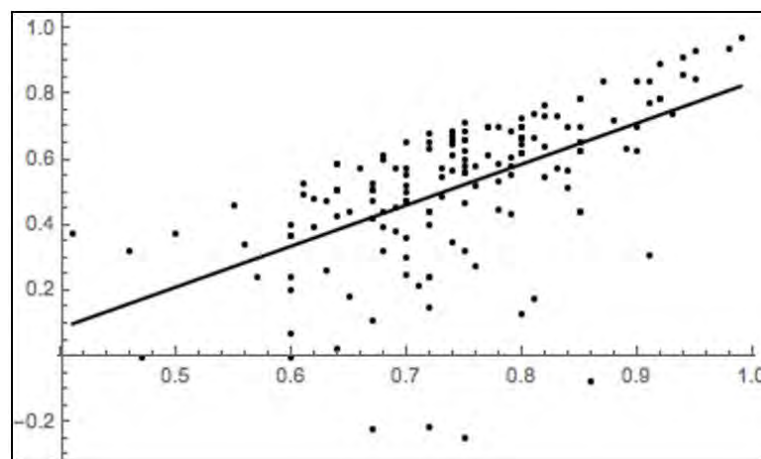


Figure 5. Individual gain versus final-test scores (N = 155 & Pearson’s $r^2 = 0.34$ for linear regression)

RESULTS

Mathematics Content Course for Pre-service Teachers

Educators of pre-service elementary teachers face a constant challenge: their students' limited understanding of what constitutes mathematics and a mathematical approach to problems. Pre-service elementary teachers are generally inflexible in their thinking, and rarely see connections between different problems or parts of mathematics. The 155 students enrolled in a pre-service teachers' mathematics course over several semesters, fell into two cohorts, cohort A with 90 students, and cohort B with 65 students, all taught by the same instructor. Primary goals of instruction for both cohorts included trying to change the severely procedural orientation to mathematics, focused on a prevailing mind-set of 'correct answers' that prospective teachers have learned to value highly and enhancing students' flexibility of thinking as they developed an increased reflective awareness of what they focus attention on as they developed schemas. Additionally, students in Cohort B were explicitly encouraged to identify and remember patterns and to establish connections, focusing on what it means to learn mathematics and on the nature of mathematics. A feature of this changed teaching direction was the explicit and intensive focus on building connections between different representations of a single problem involving binary choice in the first five weeks of the course—constructing relationships between parts of mathematics that students usually see as different. They were asked to explain how these various assigned problems were all related and were challenged to clearly state how and why the various problems were connected in ways that make sense to others, friends and sceptics alike. Opportunities for making connections with this early work were provided throughout the semester.

Students with High or Low Individual Gain

There are at least two ways we can specify what constitutes a high individual gain. The first is to use Hake's assessment of the mean $\langle \text{gain} \rangle$ as high if it is greater than 0.7 (Hake, 1988), and to say, correspondingly, that an individual gain is "high" if it is above this value. The other is a normalization in terms of a comparison group of students and say that a student's individual gain is "high", in that group, if it is more than one standard deviation above the mean of individual gains for the group.

We separated out, from the 65 students in Cohort B of 155 pre-service elementary teachers, those with gain z-score above 1 ("very high gain") or below -1 ("very low gain"). What is striking and notable is the difference in attitude to mathematics, and to learning and teaching mathematics, between the very high gain students and the very low gain students. In particular, we note a significant difference in the *engagement* of the very high gain group of students versus the very low gain group. Engagement is often conceptualized as a multidimensional concept with three main mutually dependent parts: behavioural engagement, emotional engagement, and cognitive engagement (Fredricks et al., 2004, Fredricks & McColskey, 2012; Helme & Clarke, 2001). In broad outline, *behavioural engagement* deals with effort, persistence, concentration, attention, asking questions, and contributing to class discussion, *emotional engagement* with interest, boredom, happiness, sadness, anxiety, and *cognitive engagement* with psychological investment in learning, desire to go beyond requirements, preference for challenge, flexibility in problem solving, positive coping in the face of failure, desire to comprehend and master relevant skills.

Very high gain

There were 8 students in this group (12.3% of the 65 students in Cohort B), with a mean initial-test score 0.49, mean final-test score 0.94, and mean gain 0.87. Students in this group, like most of the cohort, characterized their prior mathematics learning as instrumental (Skemp, 1976):

"I have never been taught a math course by relational understanding. All of my classes were learning rules and applying them."

"I think most of my learning in math was done instrumentally. We were taught the rules and how to use them."

"I was taught 'how' but not 'why.'"

They stated explicitly that they focused in this course on re-learning basic mathematics:

"I had to re-learn basic math in order to eventually teach it to children."

"Relearning how to count by using another system opened my mind to different ways of seeing the problems of adding."

"I am extremely grateful to have been given the opportunity to relearn this content to gain a secure foundation of mathematics."

They consistently looked for relationships and connections, wanting to understand why as well as how. Linkages o their existing framework of various mathematical concepts and processes are documented in their written work:

"A lot of the mathematics we learned has connections to something else we learned. I definitely approach math differently than I used to in high school. I now know why I use a particular method or formula.... At first glance of the locker problem, we immediately noticed that it was an issue of two choices: Open or Closed. These two choices gave us the idea of squaring each natural number. By using the square of the natural numbers 1-32 (ex: $1^2 = 1$; $2^2 = 4$; $3^2 = 9$), we figured out which lockers would remain open after the 1000th student has passed through. To test our hypothesis, we graphed the locker

numbers 1–30 and established a pattern among the lockers that remained open. To the extent of our work, we noticed that this problem is similar to that of the delivery blocks.”

“I found that I was making connections I had not before. These connections made it easier to understand what and why we were doing things in class. This influenced my attitude to change for the better. Now I’m more willing to learn new concepts and apply them to mathematics.... This assignment really opened my eyes to see how in mathematics there are patterns and how these patterns lead to connections that link them to one another.... Our work led us to the conclusion that we were working with two choices: Up/down, black/white, or right/left. All of these exercises dealt with binomials, powers of two and algebraic expansion.”

“A lot of the mathematics we learned has connections to something else we learned. I definitely approach math differently than I used to in high school. I now know why I use a particular method or formula.... The Pascalini pizza problem has the option of eight toppings. This situation is similar to the former exercises of building towers, the committee vote exercise, the grid walk problem, and the tunnel exercise with Mork. The “with” or “without” question resembles the two color combination for the tower building exercises, the “yes” or “no” vote of the committee members, the “up” or “right” direction for the grid walk and the alternating pattern of the tunnels. With these experiences in mind, we broke down the Pascalini’s problem using the algebraic expansion. The “with” or “without” strongly indicated powers of two as in the tower exercise. We extrapolated this to apply to the Pascalini’s dilemma, so we figured that $2^8 = 256$, therefore, there are 256 combinations for pizza made of 8 toppings.”

“I have learned that mathematics is indeed a series of interrelated ideas.... An example of a Linear equation can be expressed through both the Even (2,4,6,8...) and Odd (1,3,5,7...) Number Sequences. These sequences share a common bond of containing multiples of two in their own respective formulas. The Even Number Sequence general formula is represented as $2(n)$, whereas the Odd Number Sequence general formula is represented as $2n - 1$. An example of a Quadratic equation exists in the Sequence of Squares. The Sequence of Squares formula is n^2 —notice that the base changes but the exponent does not. The exponential or growth ratio also expresses a relationship as it relates to mathematical sequences. An example of the exponential or growth ratio is shown in the Doubling Sequence formula of 2^x . In this formula the base stays the same yet the exponent changes. Building Towers I and II used this particular formula to determine how many possibilities of towers there were for each individual problem.”

They emphasized the importance of being systematic in approaching mathematical problems and focused explicitly on organizational skills. They stressed organization, effort, and willingness to learn from mistakes:

“I believe my organization skills have improved, ... Do I know what to do, and why I should do it? This is what I ask myself with each assignment. Organization, effort, and willingness to learn from your mistakes are the way to truly learn math.”

“The connections have also helped my organization. I couldn’t organize my thoughts. It was like I knew what I meant, but I couldn’t explain it. ... My thoughts have become clearer ever since I’ve made better connections.”

Principally, these students became more reflective problem solvers. They were willing to reflect on their own learning, contrary to their prior mathematics experiences, and were able to elaborate what they did and did not know in very specific detail. Their ability to think more flexibly developed and they were able to switch from a direct to a reverse train of thought. Students in this group were able to see a problem and think of different ways to solve it: they focused on what the problem was asking. They focused on truly understanding a problem and being able to solve it in an efficient and elegant way and they utilized and understood appropriate mathematical terminology. This group tended not to over-generalize and were aware of what is appropriate to use in a given situation.

Very low gain

There were 11 students in this group (16.9% of cohort B), with a mean initial-test score 0.48, mean final- test score 0.64, and mean gain 0.28. This group of students split naturally into three subgroups – Group 1: 2 students; Group 2: 4 students; Group 3: 5 students.

Group 1. (Initial test z-score > 2, gain z-score < -1). There were two students in this group, with mean initial-test score 0.87, mean final-test score 0.91, and mean gain 0.22. These two students were computationally competent. Though both believed that, as teachers they need to understand how students think, they saw teaching as instruction.

“When children are given only a process and not a true explanation of material, it is the children who will suffer. Each child has a different learning process regarding mathematics and it is the job of the teacher to recognize these different methods in order to help the child understand.”

“I learned to identify how I thought by going over the series of algorithms that we went over in class, and the commutative, associative, and distributive ways to work problems. The manipulatives I learned with in class I can now use them to help others, or use them as drawings to explain with a visual what I mean.”

Like the students with very high gain, these two very low gain students were able to clearly articulate what they did and did not know. They only occasionally justified their results and were able to generalize their work. In their writings, neither student

acknowledged the importance of flexible thinking, the role of definitions or of proof. Both students focused on filling in gaps of their knowledge of procedures and on learning multiple algorithms and alternative procedures rather than on learning to think more flexibly and relationally, which was a significant focus of instruction.

Group 2. This group of four students had mean initial-test score 0.59, mean final-test score 0.68, and mean gain 0.23. Students in this group, like most students in the cohort, began the course with a very procedural approach to mathematics. Unlike the very high gain group they did not break out of this procedural approach to mathematics. They differed from students in the third group (Group 3), however, in that, like the two students in Group 1, when they were asked to use a procedure they knew, they could work a problem correctly. Three of the four students described themselves as visual learners, by which they meant: “show me how to do it.” They struggled with learning to think more flexibly and relationally, remaining focused on procedures:

“When I began this class I had a lot of trouble trying to do the mathematics relationally. It’s not all about finding the answer as I first thought when I started this class. It’s about understanding the rules behind the mathematics and providing the correct algorithm or model to show for it.”

“A number that is a factor will divide evenly into a given number.”

“I need to develop skills of having more flexible thinking. My goal is to be more comfortable with doing something numerous ways.”

All four students claimed to have become more flexible and relational thinkers yet were unable to recognize isomorphic problems or to generalize a pattern.

“In the pizza problem, we couldn’t see that there were only 2 choices, either with or without the toppings. Even though I had written (in a letter) “2 choices” was a key to solve many of the problems, I still missed it.”

Group 3. (Initial test z-score < -1, gain z-score < -1). This group of five students had mean initial-test score 0.32, mean final-test score 0.53, and mean gain 0.31. All students in this group expressed confidence in their ability to do mathematics at the end of the course. There was, however, a marked disconnect between what these students thought their understanding was, and what their instructor thought it was. For example, in the final examination a student in this group rated themselves as “Exemplary (all the time) 5/5” in creating a general rule or formula, despite writing consistently throughout the semester that they had trouble coming up with an equation. These four students characterized themselves as hands-on and visual learners and claimed to have problems with oral or written explanations. Their expressed view of being a visual learner meant seeing a problem worked on the board, not thinking in visual images. They all expressed a belief that learning mathematics is about a teacher showing how to do a problem. Then, and only then, they said, could they understand what was done. They persisted with inappropriate word usage.

Despite their belief that they were becoming more flexible thinkers, what it means to learn mathematics and to teach mathematics remained instrumental for these students. Their focus of attention was on learning how to do a procedure. They held to working one way —the way with which they were most comfortable. These students did not use multiple representations to solve problems, believing that being shown more than one way to do a problem is confusing. For example, on the final test, a student in this group was unable to demonstrate more than one way to compute subtraction problems using whole numbers and mixed numbers and was unable to divide mixed numbers correctly at all: The problem asked students to use (a) missing factor; (b) “you don’t have to multiply”, and (c) standard algorithm. Given a shaded array, this student was unable to identify the fraction multiplication problem indicated by the drawing.

Both groups of students exhibited positive behavioural and emotional engagement in that most students, in both groups, made references emotionally and behaviourally to wanting to change – in realizing there was a broader deeper view of mathematics than what they had previously experienced and in wanting to become more flexible in mathematical problem solving. However, despite the students in the very low gain group asserting they had become more flexible problem solvers, the overwhelming evidence from their efforts was that they had not. The very high gain students by way of contrast had distinct positive gains in terms of cognitive engagement, exhibiting psychological investment in learning, a desire to go beyond requirements, flexibility in problem solving, and a desire to comprehend and master relevant skills. What we find with this group of 65 students is that, when disaggregated into very high and very low gain groups, the very high gain group showed a marked level of positive cognitive engagement, while the very low gain group did not.

Comparison of The Distribution of Gain Score for Pre-Service Teachers

The distribution of gain scores for Cohort A and Cohort B prior to and after the change in focus of the changed teaching direction for cohort B were examined and depicted in **Table 2**.

Table 2. Comparison of gain scores for pre-service course pre- & post-change of instructional focus

Cohort	N	Gain			
		Min	Max	Mean	95% CI for mean
A: Change in focus	90	-0.25	0.93	0.48	(0.4, 0.54)
B: Change in focus	65	0.13	0.97	0.57	(0.52, 0.61)

The difference in mean gains is significant at the 95% level. A t-test for difference in mean gains was inappropriate due to significant departure from normality in the distribution of gains for the pre-change cohort, so a bootstrap method was used to estimate a p -value $p < 0.034$, a probability slightly greater than that of obtaining 5 heads in a row from a fair coin toss. Cohen's d is 0.36, indicating a small effect size: 58% of students in the later cohort obtained an individual gain greater than the median gain for the earlier cohort. **Figure 6** shows smooth kernel approximations to distribution of individual gains.

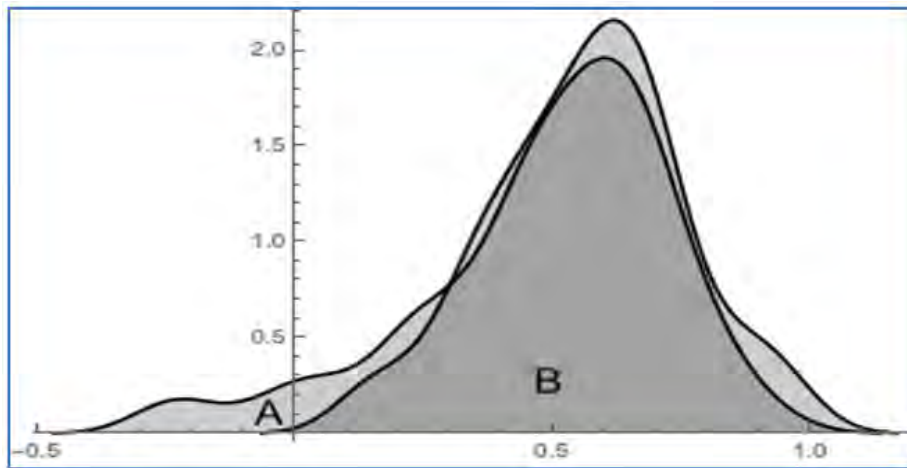


Figure 6. Smooth kernel approximations to distribution of individual gains: pre-change (A) and post-change (B)

Comparisons of Distributions of Gain Scores for Two-Year College Developmental Algebra Students

The gain scores of two-year college students enrolled in sections of a pilot developmental algebra course at a large suburban two-year college were also examined and depicted in **Table 3** and **Figure 7**. Three instructors volunteered to teach pilot sections of intermediate algebra, along with a co-author of the course textbook based on a process definition of function. The sequence of topics focused on making sense of notation, investigating problems using a variety of representations, and discovery of connections. The distribution of individual gains of the classes taught by the co-author of the text compared with the gains of the students taught by the three instructors who had no prior experience with the pilot text are described below.

Table 3. Comparison of gain scores for text co-author versus three volunteer instructors

Cohort	N	Gain			
		Min	Max	Mean	95% CI for mean
Text co-author	21	0.07	1.00	0.64	(0.53, 0.76)
Other instructors	66	-0.20	0.88	0.44	(0.38, 0.50)

The difference in mean gains is significant at the 99% level. A t-test for difference in mean gains was inappropriate due to gross departure from normality in the distribution of gains for the text author cohort, so a bootstrap method was used to estimate a p -value $p < 0.0022$, a probability approximately that of obtaining 9 heads in a row from a fair coin toss. Cohen's d is 0.83, indicating a large effect size: 76% of students in the text author cohort obtained an individual gain greater than the median gain for the other three instructors' cohort.

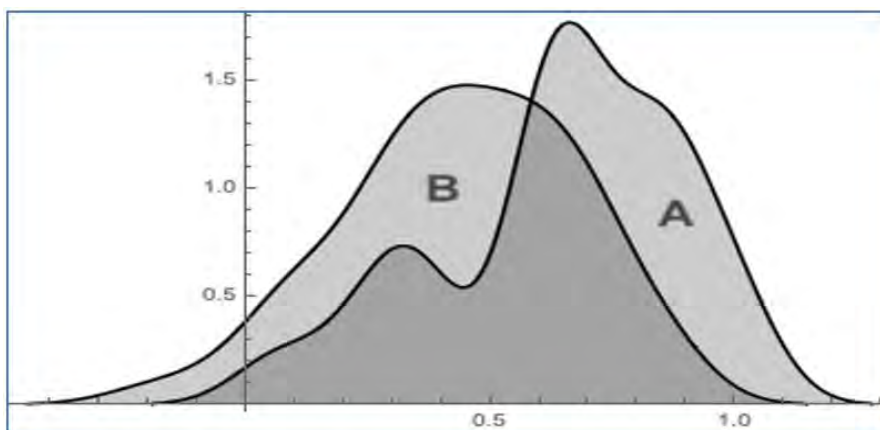


Figure 7. Smooth kernel approximations to distribution of individual gains: Text co-author, right (A) and 3 other instructors, left (B)

One instructor, two different courses

Examples of the distributions of gains for two different classes, each taught by the same teacher, three years apart are provided in **Table 4**. The actual distributions of student gains, above and beyond the simple means of the gains, or the Hake mean gain, allows us to draw statistical inferences.

Example 1. Intermediate and Beginning Algebra: Two developmental algebra courses (Cohort A: Intermediate, and Cohort B: Beginning) were taught by the same instructor. The instructor, who taught the pilot Intermediate Algebra text, used a traditional Beginning Algebra text three years later.

Table 4. Comparison of gain scores for developmental algebra courses pre- & post-change

Cohort	N	Min	Max	Gain	
				Mean	95% CI for mean
A: Intermediate	30	-0.08	0.88	0.50	(0.42, 0.59)
B: Beginning Algebra	42	-0.08	0.79	0.21	(0.11, 0.32)

A marked difference in the distribution of gains is shown by the smooth kernel approximations to the distributions of individual student gains for cohorts A and B (shown in **Figure 8**).

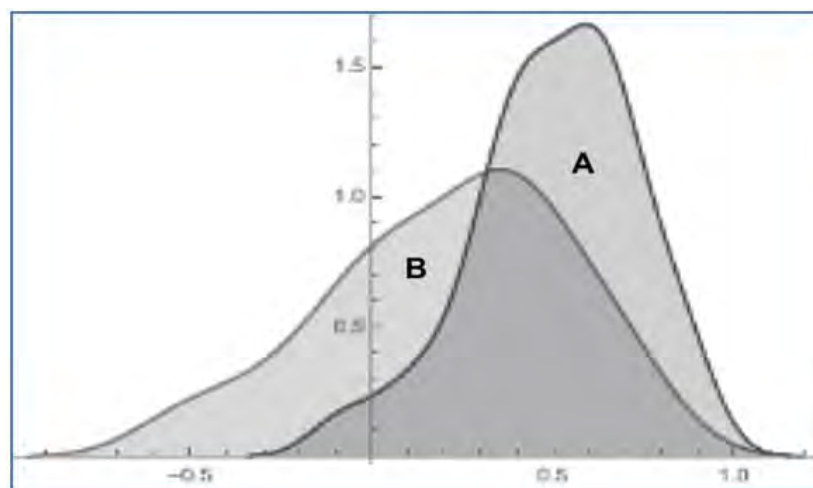


Figure 8. Smooth kernel approximations for distribution of individual student gains for cohorts A and B

The difference in mean gains is significant at the 99% level. A *t*-test for difference in mean gains returned a *p*-value < 0.000005, a probability less than that of obtaining 14 heads in a row from a fair coin toss. Cohen's *d* is 0.99, indicating a large effect size: 90% of students in cohort A obtained an individual gain greater than the median gain for cohort B.

Example 2. Intermediate Algebra and Pre-service Content Course: The text co-author taught both the pilot intermediate algebra course and, three years later, the Cohort B pre-service teacher mathematics content course. Comparison of the gain scores for the pre-service teacher course shows a marked difference in the distribution of gains as depicted in **Table 5**.

Table 5. Comparison of gain scores for Intermediate Algebra (A) and Pre-service (B) cohorts

Cohort	N	Min	Max	Gain	
				Mean	95% CI for mean
A: Intermediate Algebra	26	-0.2	1.0	0.36	(0.26, 0.47)
B: Pre-Service Content	65	0.13	0.97	0.57	(0.63, 0.61)

A marked difference in the distribution of gains is shown by the smooth kernel approximations to the distributions of individual student gains for cohorts A and B in **Figure 9**.

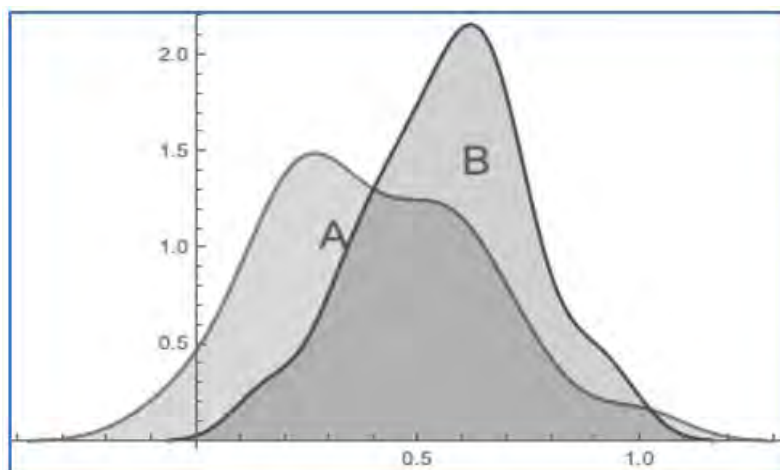


Figure 9. Smooth kernel approximations for distribution of individual student gains for cohorts A and B

The difference in mean gains is significant at the 99% level. A t -test for difference in mean gains returned a p -value < 0.0002 , a probability less than that of obtaining 12 heads in a row from a fair coin toss. Cohen's d is 0.769, bordering on a large effect size: 86% of students in cohort B obtained an individual gain greater than the median gain for cohort A. In each of these two cases of an instructor teaching two different cohorts, we see a marked and statistically highly significant difference in the distribution of gains.

DISCUSSION AND CONCLUSIONS

Utility of The Gain Statistic

The individual gain statistic, calculated for each student from an initial-test score and a final-test score, can, if the two tests are related to each other and to the course focus, provide a numerical indication of a student's cognitive engagement with the goals and aims of the course and the extent to which a student was prepared to work toward those goals. The gain function provides *significant extra* statistical information beyond initial-test scores, bears no necessary relation to the initial test score, and generally has very low correlation with it.

The individual gain, initial-test to final-test, has some desirable features that assist a teacher of mathematics to:

- estimate the overall growth and learning that took place initial to final test,
- disaggregate data to determine students with very high or very low growth, and match that to other indicators, including indicators of cognitive engagement,
- compare distribution of individual gains across different classes, and
- examine the effectiveness of one's own instructional goals and assessments of students.

The distribution of individual gains for a class or cohort of students allows a teacher to sort students into very low gain (gain z -score < -1), low gain (below average gain), high gain (above average gain) and very high gain (gain z -score > 1) to assist in relating these four groups to possible degrees of cognitive engagement with course goals.

When an instructor's course goals and teaching are aligned to what is tested in initial and final tests we expect there will, in general, be relatively high mean gains. We expect to see higher overall gains when an instructor makes their course aims explicit, teaches to those aims, and utilizes classroom techniques to enhance student cognitive engagement, since higher individual gains appear to be related to higher student cognitive engagement with course goals, instructor aims, and general overall course cognitive engagement. Conversely, we expect to see lower overall individual gains when teaching goals are unaligned with initial and final tests or when instructional methods and content presentation result in a lack of student cognitive engagement.

Other instructional benefits of using the gain statistic to examine distribution of individual student gains, include:

- Pre-tests provide a formative assessment at beginning of instruction, providing an instructor with data for specific, targeted remediation.
- For a given cohort of students, students' learning and growth, indicated by the individual gain, informs an instructor of the effectiveness of instruction for students in that cohort.
- Reflective examination of the alignment of assessment with course goals and clarification of articulated instructional goals which can result in improvement of instructional effectiveness and greater student success.

Statistical Inference Between Cohorts

The evidence we have indicates that, for a given teacher, distribution of gains may be markedly different for different classes, potentially indicating differing degrees of student engagement with course goals. For an administrator, such as a course coordinator or Department Chair, the comparison of distributions of gains across a range of instructors for a course can be revealing in terms of student growth and matching instructor course goals to student learning.

Examination of the extent to which the instructor was successful suggests that changing one's instructional practices to more actively engage students in their learning requires developing more reflective teaching practices, including thinking more deeply about one's instructional goals, the purposes and methods of assessing what students know and remember, and a willingness to examine the effectiveness of one's own teaching. We might attribute the reported marked difference in the distribution of gains to differing instructional approaches and course goals and aims of the instructors. More likely however, in our view, is that the difference in the distribution of gains, in each case, was influenced by the extent of engagement of the students in the different cohorts, and the possible styles or levels of engagement of which they were capable.

Design of Initial and Final Tests

Interpreting initial-test to final-test gains requires consideration of the purpose for which the initial-test and final-test are given, together with the alignment of initial and final-test questions with the instructional goals. In our analysis, when the pre-test and post-test were aligned with instructional goals of enhancing pre-service elementary teachers' flexibility of thinking, increasing their reflective awareness of what they focus attention on as they develop schemas and, for cohort B, the addition of an explicit and intensive focus on seeing connections between different exercise, problems and parts of the course, the pre-test and final-test are used to determine individual gain.

In comparisons where pre-tests and post-tests are not aligned precisely with instructional goals and strategies, an initial-test to post-test comparison is not a valid measure of determining individual gains. For example, all 155 pre-service teachers were administered a pre-test of 5th grade basic arithmetic skills, on which the department required a score of 80% to receive a grade of C or better for the course. Students were allowed two additional attempts to achieve the 80% score if they did not obtain it initially. The post-test was the score of this 5th grade arithmetic skills test that a student achieved when the 80% threshold was met. In this instance, individual gain was determined by the basic arithmetic skills pre-test (initial-test) and a final-test which included questions that tested conceptual understanding, problem-solving, making connections, and flexible thinking, as well as arithmetic skills included on the pre-test.

In the developmental intermediate algebra course designed for more conceptual understanding, a departmental final exam, focused on traditional skill proficiency, was required in all sections of the course. In this instance, a post-test identical to the formative assessment pre-test aligned with an instructor's goals was averaged with the departmental final exam (multiple choice) and course open response final exam scores. Gain, for this course, was determined using the initial-test score (pre-test) and the averaged final score.

Comparison of individual gains calculated pre-test to post test, and pre-test to final-test scores of the developmental intermediate algebra course show two quite different distributions of gains in **Table 6** and **Figure 10**.

Table 6. Developmental intermediate algebra (N = 26) mean of individual gains calculated from pre-test to post-test and pre-test to final-test scores

How gain was calculated	N	Gain	
		Mean	95% CI for mean
Pre-test to post-test	26	0.35	(0.25, 0.45)
Pre-test to final-test score	26	0.53	(0.46, 0.60)

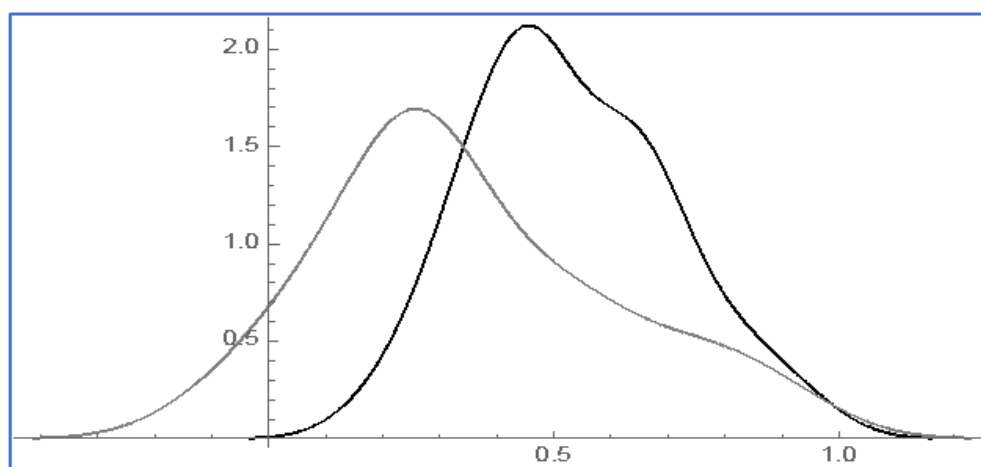


Figure 10. Smooth kernel approximations to distribution of individual gain: pre-test to post-test (left) and pre-test to final-test (right)

The moral, and the central issue to keep in mind when calculating the distribution of gain scores for a cohort of students, is that these gain figures only make sense in relation to the teaching and engagement in a course if the initial-test and final-test are connected in some meaningful way to that teaching and engagement.

Effect Size for Different Teaching Practices and Different Cohorts

One of the most compelling reason for calculating individual gains, over simply the summary mean gain, for a class or cohort, is the feature the full distribution of gains has to estimate an effect size when comparing two different teaching approaches, or a single teacher across different classes. Imagine two different student cohorts, A and B, taught by two different teachers, using two different teaching foci, for the same topic or syllabus, with the same initial-test and final-test administered to both cohorts. Utilizing individual student gains, initial-test to final-test we can calculate the percentage of students in cohort B whose gain was greater than the median gain for cohort A (or vice-versa). With no effect for the teaching experience for cohort B over that for A we expect to see just 50% of cohort B with a gain greater than the median for cohort A. The greater that percentage is over 50%, the greater the estimate of the effect. We can, of course, use such tried (and often difficult to interpret) methods such as calculating Cohen's d , or other related numerical indicators of effect size, but often what their interpretation boils down to is just what we have stated directly: the percentage of students in cohort B whose gain was greater than the median gain for cohort A (or vice-versa).

Conversely, if our argument, that individual gain indicates student growth in a class over a semester, has merit, and if a teacher adopts as close to the same teaching focus and practice for two different cohorts of students, A and B, then the percentage of individual gains for cohort B greater than the median gain for cohort A (or vice versa) is an indicator of the extent to which one cohort demonstrated greater growth in the course than did the other. In both scenarios, the distributions of gains allow us to infer whether or not the effect is statistically significant.

Author contributions: All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Declaration of interest: No conflict of interest is declared by authors.

REFERENCES

- Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, 74(10), 917-922. <https://doi.org/10.1119/1.2213632>
- Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. CRC Press. <https://doi.org/10.1201/9781420035926>
- Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukias, A., & Vincke, P. (2000). *Evaluation and decision models: A critical perspective* (Vol. 32). Springer Science & Business Media.
- Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., & Vincke, P. (2006). *Evaluation and decision models with multiple criteria: Stepping stones for the analyst* (Vol. 86). Springer Science & Business Media. <https://doi.org/10.1007/978-1-4615-1593-7>
- Brookhart, S. M. (1999). *The art and science of classroom assessment: The missing part of pedagogy*. The George Washington University Graduate School of Education and Human Development.
- Burkholder, L. (2015). Impartial grading revisited. *Teaching Philosophy*, 38(3), 261-272. <https://doi.org/10.5840/teachphil201581136>
- Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio Books.
- Chi, M. T., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wiley, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, 42(6), 1777-1832. <https://doi.org/10.1111/cogs.12626>
- Close, D. (2009). Fair grades. *Teaching Philosophy*, 32(4), 361-398. <https://doi.org/10.5840/teachphil200932439>
- Coburn, W. W., Schuster, D., Adams, B., Applegate, B., Skjold, B., Undreiu, A., Loving, C. C., & Gobert, J. D. (2010). Experimental comparison of inquiry and direct instruction in science. *Research in Science & Technological Education*, 28(1), 81-96. <https://doi.org/10.1080/02635140903513599>
- Concetta Capizzo, M., Nuzzo, S., & Zarcone, M. (2006). The impact of the pre-instructional cognitive profile on learning gain and final exam of physics courses: A case study. *European Journal of Engineering Education*, 31(6), 717-727. <https://doi.org/10.1080/03043790600911811>
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481. <https://doi.org/10.3102/00346543058004438>
- Davis, G. E., & McGowen, M. A. (2001). Jennifer's journey: Seeing and remembering mathematical connections in a pre-service elementary teachers' course. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the International Group for the Psychology of Mathematics Education* (vol. 2, pp. 305-312), Freudenthal Institute, Utrecht University, The Netherlands.
- Davis, G. E., & McGowen, M. A. (2004). Individual gain and engagement with mathematical understanding. In *Proceedings of the XVI Annual Meeting, North American Chapter of International Group for the Psychology of Mathematics Education* (vol. 1, pp. 333-341). Toronto, Canada.
- Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, 20(2), 159-165.

- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research*, 10(2), 020119. <https://doi.org/10.1103/PhysRevSTPER.10.020119>
- Fadaei, A. S. (2019). Comparing two results: Hake Gain and Dellow Gain, to analyze FCI data in active learning process. *US-China Education Review*, 9(1), 31-39. <https://doi.org/10.17265/2161-623X/2019.01.003>
- Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique*, 25(3), 175-203. <https://doi.org/10.1177/073724770002500301>
- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education*, 41(2), 121-139. <https://doi.org/10.1177/00224669070410020101>
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763-782). Springer. https://doi.org/10.1007/978-1-4614-2018-7_37
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Fuchs, L. S., & Fuchs, D. (1986). Linking assessment to instructional intervention: An overview. *School Psychology Review*, 15(3), 318-323. <https://doi.org/10.1080/02796015.1986.12085236>
- Hake, R. (2002). Lessons from the physics education reform effort. *Conservation Ecology*, 5(2), 28. <https://doi.org/10.5751/ES-00286-050228>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74. <https://doi.org/10.1119/1.18809>
- Hake, R. R. (1999). *Analyzing change/gain scores*. <http://www.physics.indiana.edu/~sdi/AnalyzingChange-Gain.pdf>
- Hammons, J. O., & Barnsley, J. R. (1992). Everything you need to know about developing a grading plan for your course (well, almost). *Journal on Excellence in College Teaching*, 3, 51-68.
- Helme, S., & Clarke, D. (2001). Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal*, 13(2), 133-153. <https://doi.org/10.1007/BF03217103>
- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, 12(4), 645-654. <https://doi.org/10.1187/cbe.13-05-0090>
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(1), 87-91. <https://doi.org/10.1119/1.2372468>
- McCrickerd, J. (2012). What can be fairly factored into final grades? *Teaching Philosophy*, 35(3), 275-291. <https://doi.org/10.5840/teachphil201235329>
- McGowen, M. A., & Davis, G. E. (2001). Changing elementary teachers' attitudes to algebra. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), *Proceedings of the 12th ICMI Study on The Future of the Teaching and Learning of Algebra* (vol. 2, pp. 438-335), The University of Melbourne, Australia. <http://repository.unimelb.edu.au/10187/2812>
- McGowen, M. A., & Davis, G. E. (2002). Growth and development of pre-service elementary teachers' mathematical knowledge. In D. S. Mewborn (Ed.), *Proceedings of the XIX Annual Meeting, North American Chapter of International Group for the Psychology of Mathematics Education* (vol. 3, pp. 1135-1144). Athens, Georgia. <https://eric.ed.gov/?q=ED471747>
- McGowen, M. A., & Davis, G. E. (2014). *Individual gain and engagement with mathematical understanding*. <https://arxiv.org/pdf/1406.2269.pdf>
- Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades*. Jossey-Bass.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research, and Evaluation*, 7(1), 3.
- Moskal, B. M. (2002). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research, and Evaluation*, 8(1), 14.
- Redish, E. F., Saul, J. M., & Steinberg, R. N. (1997). On the effectiveness of active-engagement microcomputer-based laboratories. *American Journal of Physics*, 65(1), 45-54. <https://doi.org/10.1119/1.18498>
- Setiawan, A. R., & Kudus, M. (2020). What is the best way to analyze pre-post data? *EdArXiv*. <https://doi.org/10.35542/osf.io/h4e6q>
- Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 153-166.
- Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE—Life Sciences Education*, 10(1), 55-63. <https://doi.org/10.1187/cbe.10-08-0101>
- Törnqvist, L., Vartia, P., & Vartia, Y. O. (1985). How should relative changes be measured? *The American Statistician*, 39(1), 43-46. <https://doi.org/10.1080/00031305.1985.10479385>
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37. <https://doi.org/10.1111/j.1745-3984.2002.tb01133.x>

APPENDIX A

Relation of Individual Gain Statistic to Other Relative Change Functions

The individual gain statistic is related to a class of relative change functions (Bonate, 2000, p. 75-90; Dimitrov & Rumrill Jr., 2003; Tornqvist et al., 1985). A *change function* in the sense of Tornqvist, Vartia & Vartia, is a function C of two non-negative real variables x (initial-test score) and y (final-test score) with the following properties:

- (1) $C(x, y) = 0$ when $y = x$
- (2) $C(x, y) > 0$ when $y > x$
- (3) $C(x, y) < 0$ when $y < x$
- (4) For all $\lambda > 0$, $C(\lambda x, \lambda y) = C(x, y)$
- (5) For each x , the function $y \rightarrow C(x, y)$ is continuous and increasing

For example, the proportional change score $C(x, y) = (y-x)/x$ (Bonate, 2000) clearly has properties (1) – (5) above. In contrast the individual gain $g(x, y) = (y-x)/(1-x)$, where x and y are normalized so as to lie between 0 and 1, satisfies (1) – (4), but trivially fails to satisfy (5). The proportional change function can be written as $C(x, y) = y/x - 1$ and so, in common with other change functions, can be expressed as a function of y/x . The gain function cannot be so expressed, due of course to the normalization of the test scores in calculating the gain.

The gain function is characterized by its properties in relation to the binary operation

$$x^*y := x + y - xy$$

namely, as one can easily verify, $g(x, y) = (y-x)/(1-x)$ is the unique function $g: [0, 1] \times [0, 1] \rightarrow (-\infty, 1]$ satisfying:

- (6) $g(x, x) = 0$ for all $0 \leq x < 1$
- (7) $g(0, y) = y$ for all $0 \leq y \leq 1$
- (8) $g(x, z) = g(x, y)^*g(y, z) = g(x, y) + g(y, z) - g(x, y) \times g(y, z)$ for all $0 \leq x, y, z < 1$

This feature of the gain function places it more clearly in perspective with the logarithmic difference function $L(x, y) = \log(y/x)$ which is the unique relative change function satisfying the additivity property $L(x, z) = L(x, y) + L(y, z)$ (Tornqvist et al., 1985). Because of property (8)—reminiscent of a measure in the sense of measure theory—we interpret individual gain as a numerical indicator of the ‘size’ of change from one test to a succeeding test. The gain function, therefore, is of theoretical interest as the unique measure of relative change satisfying (6)–(8) above, and a statistic that, as we documented, has generally low correlation with initial-test scores.