Individual gain and engagement with mathematical understanding

Article · June 2014

Source: arXiv

2 authors:

M. Mcgowen
William Rainey Harper College, United States

Gary Davis
University of Massachusetts Dartmouth

INDIVIDUAL GAIN AND ENGAGEMENT WITH TEACHING GOALS[1]

Gary E. Davis
University of Massachusetts-Dartmouth
Dartmouth, MA, USA
gedavis@wsu.edu

Mercedes A. McGowen
William Rainey Harper College
Palatine, Illinois, USA
mmcgowen@harpercollege.edu

Our aim in this paper is to relate pre-service elementary teachers' mathematics gain in test scores, initial-test to final-test, to psychological profiles, attitudes, and dispositions to learning mathematics. Our approach is "shamelessly eclectic" in the sense of Rossman & Wilson (1994), integrating both quantitative and qualitative methods to tell our story, because our story is one that tries to link numerical tests scores with psychological dispositions of students. Thus, we extend Hake's (1998) findings on average gain to individual students. The gain statistic assesses the amount individual students increase their test scores from initial-test to final-test, as a proportion of the possible increase each student. We examine the written work of students with very high gain and those with very low gain and show that these groups exhibit distinct psychological attitudes and dispositions to learning mathematics. In an appendix we examine a common belief that students with low initial-test scores have higher gains, and students with high initial-test scores have lower gains, and show this is not correct for a cohort of pre-service elementary teachers.

INTRODUCTION

As teachers of mathematics we want to know that our students have learned something from their class experiences, and we want to be able to assess what form that learning takes. Commonly, teachers will use a test at the beginning of an instructional sequence and an identical or similar test at the end. Increased scores from initial to final tests tell us little about the nature of student learning, its qualitative character, or how students exercised flexible thinking or a different and productive focus of attention in their learning. Initial to final test score comparisons simply tell us that a student's test score increased by such and such an amount. Yet in 1998, Richard Hake made an interesting discovery related to initial and final test scores. Hake (1998), in defining the average gain statistic as the amount students increase their test score on average from initial-test to final-test as a proportion of the possible average increase, found that traditional lecture courses in undergraduate physics are associated with a relatively low value for the average gain while courses with less emphasis on lectures and more on participation, are associated with relatively high average gain. Thus, Hake showed that a statistic obtained from pre/post test scores could distinguish lecturing style in undergraduate physics

In Fall 2000 we began to teach mathematics to pre-service elementary teachers differently. Starting from a model for helping students make explicit their implicit

[1] Davis, Gary E. & McGowen, Mercedes A. (2004). Individual Gain and Engagement with Mathematical Understanding. *Proceedings of the XVI Annual Meeting, North American Chapter of International Group for the Psychology of Mathematics Education*. Toronto, Canada. 1, 333-341.

memories of mathematics (Davis, Hill & Smith, 2000) we designed an introductory series of isomorphic combinatorial problems designed to assist students to see connections between apparently disparate problem situations, without their being told explicitly about those connections. At the same time we were looking for measures, or numerical indicators, of student growth and understanding of mathematics that were based on tests we gave at the beginning and end of a semester. We came up with the difference between final-test and initial-test scores, scaled to take into account the total marks a student might obtain from the initial to the final test – essentially Hake's gain, but calculated for each individual student rather than as an average over a class. The term *gain* will have a technical meaning throughout this paper, as defined in the section below on the gain statistic.

THE GAIN STATISTIC

Hake (1998) introduced the mean gain—denoted <gain>—for a class of students who were given an initial-test and a final-test in undergraduate physics:

$$\langle gain \rangle = (\text{mean final-test} - \text{mean initial-test}) / (1 - \text{mean initial-test})$$

This is a measure of what fraction, on average, students achieved of the possible marks they could achieve from initial-test to final-test. Hake studied the mean gain for classes consisting of over 6,000 undergraduate physics students in total, and concluded that generally high mean gains (0.48) were associated with classes with a focus on participation and interaction, while low mean gains (0.23) were associated with traditional lecture classes. In earlier studies involving pre-service elementary teachers as well as developmental algebra students we calculated, independent of Hake, an individual gain for each student as

$$gain = (\text{final-test} - \text{initial-test}) / (1 - \text{initial-test})$$

(Davis & McGowen, 2001; McGowen & Davis, 2001a, 2001b, 2002). Note that gain is undefined if the initial-test score is 1, a situation we have not encountered in the data reported here, nor in similar data over 7 years.

In this paper we discuss this statistic in some detail, placing it in perspective with relative change functions. The individual gain statistic is related to a class of relative change functions (Tornqvist, Vartia & Vartia, 1985; Bonate, 2000, pp. 75-90). Two examples of commonly used relative change functions are (i) the proportional change score *(final-test score – initial-test score)/initial-test score* and (ii) the logarithmic difference, *log(final-test score/initial-test score)*. Change functions are characterized by a list of relatively straightforward properties, but the individual gain fails to be a relative change function in the sense of Tornqvist, Vartia & Vartia (1985) due to the fairly trivial fact that the denominator in the definition of the gain statistic rules out a requisite scaling property of a relative change function (see Appendix B).

Hake (1999) points out that the average gain in his physics studies correlates poorly with initial-test scores, a finding that is in accord with our studies. This is in marked contrast to the logarithmic difference, log(final-test score/initial-test score), which correlates linearly for our data with initial-test scores ($r^2 = 0.83$). The percentage change, (final-test score - initial-test score)/initial-test score, correlates 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.

INITIAL-TESTS AND FINAL TESTS

Pre-tests and post-tests are commonly thought of as part of quasi-experimental design and as such are subject to numerous confounding variables that affect internal validity. (Campbell & Stanley, 1966, pp. 7-12; see Bonate, 2000, for a detailed discussion of pre-test/post-test design and analysis). A common objection to using similar, but not identical, initial and final tests is that in comparing student scores from one to the other we are trying to "compare apples and oranges." This would be a valid objection if we were asserting that an intervention was associated with a change in test score, initial to final test. But that is not our purpose in this article: our aim is to understand how we might assess growth in student mathematical development and understanding across a semester, given the usual assessment tools and practices available to a classroom teacher.

Essentially we use a statistic based on initial and final test scores to disaggregate student data in meaningful ways. Note that the nature of the test items is not relevant to the present discussion and analysis, even though the test items *are* integral to understanding the nature of the course. Our analysis is based around the different attitudes to learning and engagement exhibited by students with differing gain statistic, as obtained from a final and initial test that we, the instructors, deemed relevant to the course itself. In this paper we assume test scores have been normalized so as to lie between 0 and 1. In the application of the gain statistic, we are interested in student attitudes and dispositions to learning mathematics, rather than comparing mean test scores before and after an instructional treatment. We use different, yet strongly related, tests in a sequence—one near the beginning of a course, one nearer to the end. We use the terms "initial" and "final" test to alleviate confusion that might result from use of pre-test and post-test in these circumstances.

DATA SOURCES/EVIDENCE

We consider pooled data from 4 classes of a 16 week pre-service elementary mathematics course. The cohort for whom complete data was available consisted of 65 students. Students were given a written mathematics competency test (referred to as "initial-test") the first week of the semester. The students sat a final written examination at the end of the course. This written final examination contained problems that required students to recognize the mathematics and skills in contextual situations along with problems similar to those included on the competency test that tested skills. We refer to the final examination as "final-test". Test scores have been scaled so as to represent numbers in the range 0 through 1.

We examine the written work over a semester of those students who had gain more than one standard deviation above or below the mean cohort gain, for evidence of attitudes and dispositions to learning mathematics. Note that the students with very high gain necessarily had high final test scores because

$$\text{final-test score} = \text{gain} + [\text{initial-test score} \times (1 - \text{gain})]$$

and the second term on the right side of the equation is non-negative. Importantly, however, not all students with high final-test scores had high gain.

RESULTS

VERY HIGH GAIN

This group of students had gain statistic more than one standard deviation above the class average of 0.57. There were 8 students in this group (12.3% of the cohort), 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." LT
>
> "I think most of my learning in math was done instrumentally. We were taught the rules and how to use them." SM

However, 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." JH
>
> "We have essentially (to my mind) be re-learning mathematics." JK
>
> "I felt like I am re-learning everything." SM

They consistently looked for relationships and connections, and stated that how they approach a mathematics problem had changed:

> "My self-confidence in my ability to do mathematics has increased. Mathematics is making a lot of sense to me now. 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." JW
>
> "… 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. I'm more willing to learn new concepts and apply them to mathematics." JV

They wrote that their organizational skills had increased:

> "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." JH
>
> "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." SM
>
> "Three habits of mind which make math a lot easier to complete are: think, estimate answers, and learn to use patterns. In using these three habits, I have grown in my organizational skills as well." HH

Principally, these students had become more reflective problem solvers–a change from their prior mathematical experiences. They were able to elaborate what they did and did not know in very specific detail. 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. 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. This group tended not to over-generalize, and were aware of what is appropriate to use in a given situation. 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. They had a focus on looking for relationships—not only looking for isomorphic problem situations.

VERY LOW GAIN

This group of students had gain statistic more than one standard deviation below the class average. There were 11 students in this group (16.9% of the cohort), 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 A: 5 students; Group B: 2 students; Group C: 4 students.

Group A.

This group had mean initial-test score 0.24, mean final-test score 0.50, and mean gain 0.33. All students in this group expressed confidence at the end of the course in their ability to do mathematics. There was, however, a marked disconnect between what these students thought their understanding was, and what we thought it was. For example, in the final examination students in this group rated themselves as "Exemplary (all the time) 5/5" in creating a general rule or formula, despite their writing consistently throughout the semester that they had trouble coming up with an equation. All these students characterized themselves as hands-on and visual learners and claimed to have problems with oral or written explanations. However their expressed view of being a visual learner meant seeing a problem worked on the board, not thinking in visual images. This group of students claimed to learn better from examples. They all expressed a belief that learning mathematics is about the teacher showing how to do a problem. Then, and only then, they said, could they understand what was done. The following comments are typical:

> "I am more of a hands-on or visual type of learner when it comes to any subject. When a teacher verbally explains how to do some sort of math problem, I have a harder time grasping the concept. If a teacher shows and explains a problem on the board, I can actually understand. I can also learn better from examples. I also can teach myself many things, just by looking at an example." CL

> "I am a visual person. In order to understand a math problem, I need examples of the same type of problem. Usually, I can figure out how a teacher came to an answer just by looking at his/her example, and then I do really well on assignments. If I don't know how to do a problem, and we go over problems in class, I raise my hand and explain what I don't know. By the time the teacher finishes the problem, I feel better understanding how he/she got to it. However, if a teacher does not teach, I get lost." BK

Despite claims to the contrary, these students persisted in the belief that teaching means the teacher "shows me how to do it" and then "I can understand what was done".

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:

> "To really learn math, a person has got to have a feel for it. This can be accomplished by having specific examples and a concrete way of learning it. Algorithms give you a sure plan on how to do the problem, it helps you

understanding how you want to work it, and it gives you full directions on how it is supposed to be done. " NA

"The way I like to learn is when a teacher goes up to the blackboard or overhead projector, and demonstrates the mathematics by showing the process." BK

They persisted with inappropriate word usage. The examples below, in which the students use the inappropriate word "equation" instead of "expression", are typical:

"Finding an equation to match a pattern is a different story. CL

"So 2^n is not the correct equation." CL

Despite claims that their goal was to learn different algorithms to do a problem, these students did not use multiple representations to solve problems and stated that being shown more than one way to do a problem is confusing. They held to working one way—the way they were most comfortable. 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. This type of response was typical for this group.

Their reflections were frequently written as instructions to a third person—a teacher:

"If a teacher can give you a rule and then explain and show the children how and why it works, they believe you more." CL

"Teachers need to understand the algorithm and make sense of it. Students need rules with reasons. They need to be taught or shown different methods of finding an answer." LL

"You, the teacher, need to not only be able to solve the problem but to learn different process of how to arrive at an answer." BK

Group B.

There were two students in this group, with mean initial-test score 0.82, mean final-test score 0.86, and mean gain 0.22. These two students were computationally competent, and saw no need to re-think basic mathematics. They viewed teaching as direct instruction and focused on the importance of knowing how children think:

"I have learned how I thought about problems. By knowing how I think about problems and the different ways to think about them, I am able to see how students work their problems so I can either help them or I can learn from them. A very important aspect in teaching I believe I have learned is that you need to know how a student thinks about a problem before you tell him how to work it." AS

"This experience has taught me just how imperative it is for a teacher to be sensitive to each individual child's learning. The biggest aspect of this class that I will take with me is the idea that each child has a different learning process regarding mathematics and it is the job of the teacher to recognize these different methods in other to help the child understand." NM

Group C.

This group had mean initial-test score 0.6, mean final-test score 0.7, and mean gain 0.25. Students in this group, like the majority of 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 were different from students in group A, however, in that they did know a correct procedure to use, and when they were asked to use a procedure they knew, they could work the problem correctly. In their reflections and self-evaluations, they described what they should learn and their limitations:

> "I should be able to think flexibly once I look at math problems. I shouldn't be stuck with one rule or certain ways to solve it." KN
>
> "I need to develop skills of having more flexible thinking. I get frustrated easily when I can't figure out a problem." ME
>
> "Different algorithms should be used to work out different problems. You need to know when to use which algorithm in different problems." GY

CONCLUSIONS

Students in this study with very high and very low individual gains had markedly different psychological profiles in relation to attitudes to study and the course material.

Students with gain more than one standard deviation above the cohort mean worked hard and smart, particularly in relation to learning to be more organized. They focused on looking for relationships and re-learning basic mathematics so as to teach it better.

Students with gain more than one standard deviation below the cohort mean were not homogeneous in their attitudes and dispositions toward learning mathematics. The largest group began and remained very instrumental, highly dependent on being shown how to do worked examples. A second group comprised two students with high initial-test score and very low gain. These two students were competent in terms of mathematical computation and viewed teaching only as an instructional process. A third group comprised students with moderate initial-test scores and very low gain. These students were competent in using algorithms, but showed no flexibility in their approach to problems.

In the context of this study, very high gain meant engagement with the explicit aims and goals of the course and a willingness to take risks in learning mathematics. Although there were three identifiably different groups of students having very low gains, all these students showed a lack of engagement with the aims of the course—albeit in different ways—and also showed no evidence of risk-taking in relation to learning mathematics.

On the basis of the test scores and written work of these 65 pre-service teachers we are now inclined to see high individual gain as an indication of engagement with the aims and goals of an instructional sequence, and low individual gain as a lack of such engagement for varying reasons.

We believe that single classroom data from initial and final tests can be very useful provided one disaggregates the data and relates a statistic such as gain to attitudes to learning of individual students. The salient point for us is our hypothesis that irrespective of the initial-test and final-test questions—so long as they relate to the instructor's aims and goals for the course—we will see similar psychological profiles among the students with very high and very low gain. This hypothesis is eminently testable, and is not dependent on the precise nature of the test questions. Thus, our focus is not on the exact nature of test questions, nor on the details of an instructor's aims and goals. Rather, our focus is on whether a single change statistic—the individual gain—does consistently disaggregate students in the ways we have indicated in this study.

REFERENCES

Bonate, P.L. (2000) *Analysis of Pretest-Posttest Designs*. Chapman & Hall.

Davis, G., Hill, D. & Smith, N. (2000) A memory-based model for aspects of mathematics teaching. In T. Nakahara & M. Koyama (Eds.) *Proceedings of the 24th conference of the International Group for the Psychology of Mathematics Education*, vol 2, p.225-232. Hiroshima, Japan: Hiroshima University.

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. Utrecht: Freudenthal Institute, Utrecht University, The Netherlands.

Davis, G.E. & McGowen, M.A. (2003) Characterizing the gain function. Retrieved on December 18, 2003, from www.wsu.edu/~gedavis/gain.htm

Hake, R.R. (1998) Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*. 66, 64–74. Online at http://www.physics.indiana.edu/~sdi

Hake, R.R. (1999) Analyzing change/gain scores. http://www.physics.indiana.edu/~sdi

McGowen, M.A. and Davis, G.E. (2001a) Changing elementary teachers' attitudes to algebra. In Chick, H., Stacey, K., Vincent, J. & Vincent, J (Eds.) *Proceedings of the 12th ICMI Study on The Future of the Teaching and Learning of Algebra*, Vol. 2, 438-335. Melbourne: University of Melbourne.

McGowen, Mercedes A. and Davis, Gary E. (2001b). What mathematical knowledge do pre-service elementary teachers value and remember? In R. Speiser, C.A. Maher, & C. N. Walter (Eds.) *Proceedings of the XXIII Annual Meeting, North American Chapter of International Group for the Psychology of Mathematics Education*. Snowbird, Utah, Vol. 2, 875-884.

McGowen, Mercedes A. and Davis, Gary E. (2002). Growth and Developent of Pre-Service Elementary Teachers' Mathematical Knowledge. In Denise S. Mewborn (Ed.) *Proceedings of the XIX Annual Meeting, North American Chapter of International Group for the Psychology of Mathematics Education*. Athens, GA. Vol. 3, 1135-1144.

Rossman, G. and Wilson, B. (1994) Number and words revisited: Being shamelessly eclectic. *Quality and Quantity*, 28, 315-327.

Skemp, R.R. (1976) Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 153-166.

Tornqvist, L., Vartia, P. & Vartia, Y.O. (1985) How should relative change be measured? *American Statistician*, 39, 43-46.

APPENDIX A: GAIN & INITIAL-TEST SCORE

In this appendix we address the question of whether low initial-test score implies high, and high initial-test score implies low gain. At presentations of work on the gain statistic we often hear assertions like: "of course, students with low initial-test score will have high gain." The argument is that a student with a low initial-test score has a lot more room for improvement, and so a potentially higher gain, than a student with a high initial-test score. A natural corollary of this line of reasoning is that a student with a high initial-test score will have a generally lower gain than other students, "because it's harder to achieve a higher final-test score starting from a high initial-test score." The mean initial-test score for the cohort of 65 pre-service teachers in the present study was 0.39, the mean final test score was 0.74, and the mean gain was 0.57. The group of students (n = 36) with below average initial-test scores had a statistically significantly lower average final-test score than those students (n = 29) with above average initial-test score (mean of 0.70 as compared to a mean of 0.79; $p < 0.002$). However, the average gain for the group with below average initial-test score was not statistically significantly different from that of the group with above average initial-test scores (gain of 0.60 compared with a gain of 0.53; $p > 0.1$)

A similar situation pertains if we take the group with initial-test z-score < -0.5 on the one hand, and the group with initial-test z-score > 0.5 on the other. For this cohort, therefore, there seems to be little basis for the claim that low initial-test score entails high gain on average. Indeed, there are significant numbers of students who had below average initial-test scores and below average gains, as well as significant numbers with above average initial-test scores and above average gains: 13/65 = 20.0% of the cohort had below average initial-test score and below average gain (95% confidence interval = [12.1%, 31.3%]), and 14/65 = 21.5% of the cohort had above average initial-test score and above average gain (95% confidence interval = [13.3%, 33.0%]).

Of the 36 students with below average initial-test score, 13/36 = 36.1% had below average gain (95% confidence interval = [22.5%, 52.4%]), and of the 29 students with above average initial-test score, 14/29 = 48.3% had above average gain (95% confidence interval = [31.4%, 65.6%]). Thus, for this cohort of 65 students, given a student had a less than average initial-test score there was more than a 1 in 3 chance that the student had a less than average gain. Equally, given a student had a greater than average initial-test score there was about a 1 in 2 chance that the student had a greater than average gain. These proportions are not inconsiderable, and while it is more likely that a student with a below (*resp.* above) average initial-test score will have an above (*resp.* below) average gain, it is by no means a foregone conclusion.

APPENDIX B: RELATIVE CHANGE FUNCTIONS

A change function in the sense of Tornqvist, Vartia & Vartia (1985), is a function C of two non-negative real variables x (initial-test score) and y (final-test score) with the following properties:

$C(x, y) = 0$ when $y = x$

$C(x, y) > 0$ when $y > x$

C(x, y)< 0 when y < x

For all $\lambda > 0$, C (λx, λy) = C(x, y)

For each *x*, the function $y \rightarrow (x, y)$ is continuous and increasing

For example, the commonly used 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, in contrast, cannot be so expressed, due of course to the normalization of the test scores in calculating the gain.

Further, the gain function is characterized by its preservation of the binary operation $x*y := x+y – x\times y$, namely $g(x, y)$ is the unique function C:[0,1)×[0,1] → (-∞, 1] satisfying:

C(x, x) = 0 for all $0 \leq x < 1$

C (0, y) = y for all $0 \leq y \leq 1$

C(x, z) = C(x, y) + C(y, z) – C(x, y) × C(y, z) for all $0 \leq x, y < 1, 0 \leq z \leq 1$

One checks easily that g(x, y) has these properties and, conversely, if C is such a function then from 0 = C(x, x) = C(x, 0)*C(0, x) = C(x, 0) + C(0,x) – C(x, 0)C(0, x) = C(x,0) + x – xC(x,0) and C(x, y) = C(x,0)*C(0, y) = C(x, 0) + C(0, y) – C(x, 0)C(0, y), we see easily that C(x, y) = (y – x)/(1–x) = g(x, y) for all $0 \leq x < 1$ and $0 \leq y \leq 1$.

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 $L(x, z) = L(x, y) + L(y, z)$ (Torqvist, Vartia & Vartia, 1985). Because of formula C—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 A – C above, and a statistic that has low correlation with initial-test scores.