Using the effect sizes of subtasks to compare instructional methods: A network model

Garry Johns, Christopher Nakamura, and Curtis Grosse
Saginaw Valley State University

Abstract

Networks have become increasingly important in studying air pollution, energy use, genetics and psychology. These directed graphs also have features that may be useful in modeling student learning by answering questions such as the following: How can we determine if one teaching approach has better outcomes than a second method? In this paper we present a framework for dividing an approach into subtasks, assigning a numerical value (such as an effect size) to each subtask and then combining these values to determine an overall effectiveness rating for the original approach. This process allows researchers to investigate potential causes for student achievement rather than simple correlations, and can compare the effectiveness of a method for various types of students or instructors.

Key words: effect size, teaching methods, subtasks, network

Introduction and context

As university faculty face the challenge of teaching a new generation of students, some have adopted alternative assessments or a variety of teaching methods to enhance the learning experience (Cohen, 1977, Hastings, 2006, Hattie, 2009). In general, this transition away from lecturing has been slow as indicated by a recent survey of 700 calculus instructors in which a majority still believes that students learn best from clear and well-prepared lectures (Bressoud, 2011). In a related survey of over 700 faculty members who teach introductory physics, an impressive 72% had used at least one research-based instructional strategy; however, nearly one-third of this 72% no longer use any of the strategies (Henderson et al., 2012). Why do so many instructors think and react this way? Three major reasons are the following: (1) some may not be aware of the existing research supporting new approaches, (2) many are skeptical about the effectiveness of newer methods (often based on their own observations), and (3) if a faculty member is willing to try a different approach, which choice among the alternatives should she choose to produce the greatest impact?

The focus of this paper is to address reason (3) from above on how to select the best approaches. Issue (1), increasing faculty awareness of current research, is already being addressed by several professional organizations. The National Council of Teachers of Mathematics (NCTM) has recently published a survey of over fifty studies related to seven principles behind motivational strategies (Middleton & Jansen, 2011) and its 73rd yearbook, *Motivation and Disposition: Pathways to Learning Mathematics*. The Mathematical Association of America (MAA) sponsors the SIGMAA on RUME and strands on teaching and learning theory at its national meetings each January and August. Similarly, the American Mathematical Association of Two-Year Colleges (AMATYC) highlights research-based topics in several sessions at its annual meeting. As for issue (2), if faculty members have a way of choosing more effective methods, then their skepticism may be diminished due to better classroom results and replaced by a lasting commitment to incorporating classroom change.

Educator Spencer Kagan (Kagan & Kagan, 1998, p. xxii) claims that "the greatest sustained change results from the smallest changes in instruction;" however, a challenge related to selecting methods is that one general "method" may be accomplished in several ways – each with varying levels of success. For example, suppose two instructors wish to motivate students more. Instructor A chooses to focus on the *relevance* of mathematics since there seems to be general agreement that mathematics, particularly for low-achieving adolescent students, must be made more relevant in order to increase student performance (Haylock, 1999). However, Haylock cautions that not all applications are equally motivating to students. For instance, when the teacher states that certain material will be used in the next chapter or in the next course, students do not value the content as much as if they see the topic addressing a perceived need. However, even though many students see the task of finding the cost to carpet a 12 foot by 18 foot room when carpet costs $6.75 per square yard as a real problem, it is seen as someone else's problem and not the students'. Thus, it is not as motivating as if the students see an immediate, personal need for the content.

Similarly, Instructor B desires to motivate students; however, she chooses *social interaction* rather than relevance. At many institutions one of the major areas addressed in student evaluations of faculty has been "Student-Instructor Interaction." Contact between students and faculty, as well as student-student interactions such as reciprocity and cooperation among students are two of the seven research-based principles in undergraduate education advocated by Chickering and Gamson (1987). She considers the following passages from (Middleton & Jansen, 2011):

> Mathematics classrooms' social dimension can support students' learning of mathematics, particularly when teachers purposefully structure opportunities for social needs to converge with academic needs. All students' needs for relatedness – among them avoiding disapproval, achieving social affiliations, demonstrating competence, acquiring social concern, and building shared meaning – can become channeled into opportunities to engage in mathematics. Teachers' efforts to support students' mathematics learning – how they choose mathematical tasks, treat students' errors, evaluate students, reduce competition, raise status, and build positive relationships with students – can help students meet their needs for relatedness as well.[2]
>
> Integrating all of these practices into your instructional repertoire at once is not realistic . . . cycling them gradually into your teaching can help scaffold student learning.[3]

Observe in these two excerpts that many classroom changes are endorsed; however, are all of the revisions listed in the first passage important to the progress of adult learners, and if so, which changes should be prioritized (rather than merely cycled through) to achieve the greatest impact quickly?

A current approach for comparing teaching methods is to compute Cohen's effect size, *d*. This statistic was popularized in the 1960's by Cohen (Cohen, 1977) and has been used extensively to evaluate the effectiveness of many educational approaches on student achievement. This number is computed by finding the difference between two means and then dividing this difference by the paired standard deviation (if the data is matched) or by the pooled standard deviation (if the sets of data are independent). Two common scenarios where the effect size arises in educational studies are the following where the difference of means in the numerator is either (i) the average score on a pre-test subtracted from the average score on a post-test, or (ii) the mean score from a test for a control group subtracted from the mean score of a treatment group.

The effect size is easy to compute – even in meta-analyses of several studies with varying populations and sample sizes, and it is considered reliable. For instance, John Hattie has compiled the value of *d* from large meta-analyses for almost 150 educational interventions including the following (Hattie, 2009):

1)	Staying in college residence halls	$d = 0.05$
2)	Cooperative learning	$d = 0.41$
3)	Teacher-student relationships	$d = 0.72$
4)	Providing formative evaluation of programs to teachers	$d = 0.90$

These results can be interpreted as follows: there seems to be little – if any – change in achievement scores for students simply staying in college residence halls, while students who participate in cooperative learning experiences with other students see a greater improvement in achievement scores. However, developing relationships between the teacher and students, or providing feedback to instructors, appear to be associated with even more growth in student achievement.

Hattie's work focuses mainly on the effect sizes between various *approaches* or *tasks* and the final *outcome* of student scores – without considering which intermediate *tasks* or student *responses* may be potentially high-impact revisions.

A network model

The theoretical model described in this paper to address the questions of selection and priority divides a particular instructional method into a sequence of "subtasks." For the rest of this paper we will think of a method as a path with the following six steps:

1) instructor
2) motivational *principle*
3) *approach*
4) *task*
5) student *response*
6) *outcome*

For instance, if an instructor tries to motivate students using the *principle* of social interaction, then one path that might produce higher student scores would include the student-student interaction *approach* (rather than the student-instructor interaction approach, for example), followed by a classroom *task* of having pairs of students solve problems where the two students take turns explaining how to solve a problem to each other. This task results in the student's *response* (or attitude) such as valuing the mathematics or feeling more confident, and ultimately leads to an *outcome* such as student achievement as measured by test scores. This progression is shown by the path in Figure 1.

Instructor → Social interaction → Student-student interaction→ Explain in pairs → Value math → Test score

Figure 1. A "subtask" path linking the instructor and the student

In reality, a single motivational *principle* can be addressed with various approaches, one *approach* can be accomplished with several tasks, some *tasks* may invoke multiple responses (or two tasks may produce the same response), and several *responses* may result in the same outcome. Thus, a network models a more complete picture of the interactions between various subtasks. In Figure 2 a network is shown which includes several (but not all) subtask paths from the instructor to the measure of student achievement – the phrases will be referred to as *nodes* and the arrows called *arcs*. The paths begin with the instructor choosing one motivational *principle* from three (social interaction, technology, and immediate feedback). Moving in the direction of the existing arcs, the teacher next chooses an *approach* in the second column that aligns with the motivational principle, followed by one of several possible *tasks* in the third column that support that approach. Each task correlates with at least one student *response* in the fourth column which ultimately correlates with the desired *outcome* (student achievement). As already mentioned, the network shown is not complete, since there may be other motivational principles, approaches, tasks and responses not shown (as illustrated by the node labeled "In-class Technology" and the arc emanating from it).

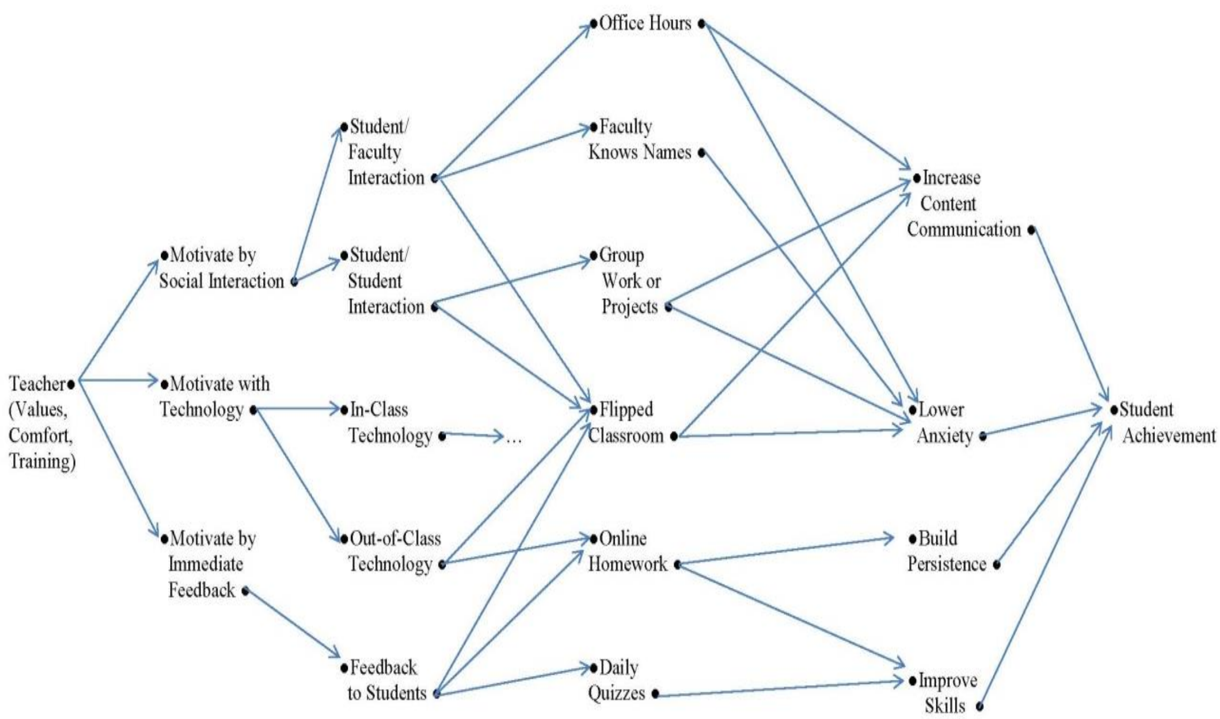

Figure 2. A network model linking the instructor and the student

Networks have become increasingly important in studying fields such as air pollution, energy use, genetics, psychology, economics, ecosystems, voting behavior, and traffic flow (Roberts, 1976). These directed graphs also have features that may be useful in modeling student learning. As is, the network in Figure 2 shows potential links between tasks and responses or between responses and outcomes. However, how can we determine if Task A results in a larger response than Task B, or similarly, when does one response lead to a better outcome than a second response? For instance, in Figure 2, if an instructor wants to provide relatively immediate feedback to her students, she could incorporate daily quizzes, online homework or a flipped classroom format. The quizzes and online homework result in just one student response each (improvement of skills and persistence, respectively) while the flipped classroom leads to both lower anxiety and increased content communication between the students. One might think that addressing two responses is better than focusing on just one in improving student achievement, but (Dweck, 2008) and (Reason, 2009) claim that persistence is seen as a necessary pre-requisite response to the outcome of student success and in (Perkins-Gough, 2013), Angela Lee Duckworth is said to argue that persistence – or "grit" – is a better indicator of success than talent or intelligence. Thus, tasks that develop the single response of persistence may be more productive than tasks that result in multiple responses.

Arc weights

By assigning a numerical value (or *weight*) to each arc, the intent is that possible comparisons could be made between various principles, approaches, tasks, or responses by combining the numbers in some way to determine an overall effectiveness. One value that has been used in similar problems in the area of path analysis is the Pearson correlation coefficient, *r* (Simpkins et al., 2006). Thus, the weights in our network could be based on the correlation determined by known statistical studies. For instance, if a strong positive association existed between the task of "having students explain problems to each other in pairs" and the response where "students value mathematics," then the correlation coefficient, *r*, would be close to 1. Unfortunately, the correlation coefficient has some limitations. First it indicates only correlation – not causation. Second, *r* measures only *linear* correlation. Third, there is no natural way to combine the correlation coefficients for the arcs on a path to determine a cumulative correlation for the entire path.

A second value that could act as the weight of an arc is the *effect size, d*, discussed earlier with Hattie's work. This parameter seems more viable because statistical studies can be done to find *d* for each of the arcs in the network, some overall values of *d* are already known (for example, $d = 0.43$ for the motivational principle of immediate feedback (Hattie, 2009)), and formulas can be developed for combining the *d*-values along a path to determine an overall effect size.

One last note about the arc weights relates to the values assigned to the arcs between the teacher and the motivational principles in the first step in the network. That is, how does one measure why an instructor chooses one principle over another? A possible method would be to use an attitudinal survey to rank the instructor's value of, comfort with, and training in that principle. It is natural to believe that if one instructor values the use of technology more than a second instructor, then the achievement for the students of the first instructor will probably be greater than that of the second instructor if both use technology in their classes; however, a second question worth studying is the following: If an instructor does not value (or is uncomfortable with) a motivational principle, could the student achievement of her students still be higher than if she focused on another principle she valued more?

Conclusion

In an effort to select more effective instructional methods, Hattie and others have used effect sizes to relate educational interventions with student achievement. This process may be able to be refined by dividing a teaching approach into subtasks and creating a network model of the interactions between these subtasks. By studying the effect of each subtask and assigning a corresponding value to each arc of the network, we may be able to determine if certain classroom tasks should be implemented, or if specific student responses should be targeted to attain more growth in student achievement. This theoretical model provides a framework for designing statistical studies that determine the arc weights and the opportunity for using methods from graph theory to combine these weights into cumulative values that may help in evaluating questions in mathematical instruction at the college level.

References

Bressoud, D. (2011). The Worst Way to Teach. *Launchings*.

Chickering, A. & Gamson Z. (1987), Seven principles for good practice in undergraduate education. *The American Association for Higher Education Bulletin*, March, 3-7.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York, NY: Routledge.

Dweck, C. (2008). *Mindset: The new psychology of success.* New York, NY: Random House.

Perkins-Gough, D. (2013). The significance of grit: A conversation with Angela Lee Duckworth. *Educational Leadership*, **71,** 14-20.

Gold, B., Keith, S. and Marion, W. (Ed.) (1999). *Assessment practices in undergraduate mathematics*. The Mathematical Association of America.

Hastings, N. (Ed.). (2006). *A fresh start for collegiate mathematics: Rethinking the courses below calculus*. The Mathematical Association of America.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Haylock, D. (1999). *Helping low achievers succeed at mathematics: Grades 2 -8* (pp. 19-21). Toronto, Ontario: Trifolium Books, Inc.

[10] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj. The use of Research-Based Instructional Strategies in Introductory Physics: Where Do Faculty Leave the Innovation-Decision Process? (2012) Accepted for publication in *Physical Review Special Topics – Physics Education Research*.

Kagan, S. and Kagan, M. (1998). *Multiple intelligences.* San Clemente, CA: Kagan Cooperative Learning.

Lenker, S. (1998) *Exemplary programs in introductory collegiate mathematics*. The Mathematical Association of America.

Middleton, J. and Jansen, A. (2011). *Motivation matters and interest counts.* Reston, VA: National Council of Teachers of Mathematics.

Reason, R. An examination of persistence research through the lens of a comprehensive conceptual framework. *Journal of College Student Development*, **50**(6), 659-682.

Robets, F. (1976). *Discrete mathematical models.* Englewood Cliffs, NJ: Prentice-Hall.

Simpkins, S., Davis-Kean, P. & Eccles, J. Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, **42**(1), 70-83.