

Game theory approaches to grading : An experiment with two incentive point systems

著者名(英)	Tim Newfields
journal or publication title	The economic review of Toyo University
volume	32
number	2
page range	33-43
year	2007-03
URL	http://id.nii.ac.jp/1060/00001708/



Game theory approaches to grading: An experiment with two incentive point systems

外国語教員の能力開発と評価リテラシー

Tim Newfields

Abstract

After highlighting a few game theory concepts, this paper describes a typical classroom in game theory contexts. Two incentive grading systems are then introduced and the merits and demerits of each system are briefly summarized. Some key points for the success of any grading system are then mentioned as well as some areas for further research.

Keywords: game theory, grading systems, educational assessment, reinforcement learning, class grading, reward incentives

要 約

本論は、評価リテラシーの最重要概念を述べた後、3つの異なる母集団に対する可能な運用方法と、評価リテラシーのための合理的な基準の開発を説いている。本論では、修正 Angoff の手法と専門家により評価が得られた3つの異なる母集団のための評価リテラシー 100項目を用いた。記述したサンプル項目と評価問題に関する70項目は、高等学校の外国語教員の評価に適したものである。本論は最後に高等学校の外国語教員のための評価リテラシーの有用性と限界、そして今後の研究の方向性について言及している。

キーワード：評価基準、展開スキル、テスト能力、統計のリテラシー、試験問題の開発

Introduction

Initially envisioned as a branch of applied mathematics and way of describing economic behavior, game theory concepts have been used in a wide variety of social disciplines (Osborne, 2004). From a game theory perspective, school classes can be seen as interactional matrices in which teachers and students try to adopt optimal behaviors in order to minimize losses and maximize returns. Though it may be somewhat simplistic to describe classes in such terms, the metaphor is useful in some respects. Since education involves a substantial in-

vestment in time, energy, and money it behooves all parties concerned to figure out what their best interactional strategy is for the conditions they confront.

A key point in game theory is that outcomes are not determined in isolation: individuals interact within the constraints of their matrices in a spirit of rational self-interest in order to make choices they believe offer maximum utility. Arguably, not all human behavior is rational, yet game theory can offer glimpses of ways teachers and school administrators can increase the likelihood of desired behaviors and reduce the likelihood of those which are not by systematically manipulating the variables inherent in each system.

After suggesting how a “typical” classroom might work in terms of game theory, two alternative classroom scenarios in which the grading variable is systematically manipulated are described. Often changing one critical variable in an interactional system will, according to game theory, influence the *likelihood* of a certain behavior manifesting. However, it should be emphasized there is seldom any certainty since all social systems have many interacting variables and a degree of randomness is a feature of the system.

A “typical” classroom game

Often classes represent an *asymmetric ultimatum game* (Foundation for Teaching Economics, 2006) in which the teacher specifies what “prices” must be paid to obtain class credit. The “price” is typically a required level of class attendance and a minimal cut-off score on a series of tests. Since the game is asymmetric, the penalties for non-compliance vary. Students who fail to “win” the game by earning credit at first are often required to repeat the game until doing so. Some drop out of the system rather than risk repeated failure. Conversely, if a teacher fails a small percentage of students for non-compliance, there are usually no penalties. Indeed, a small number of students are almost expected to fail in many schools (Covaleskie, 1994). However, a tacit game rule in many schools is that the *majority* of the students should pass regardless of whether or not actually they meet external criteria. According to game theory, both teachers and students are subtly coerced by expected outcomes. Understanding precisely what the expected outcomes are is necessary for success in playing any game.

Let us examine a typical classroom in more detail. What rewards are there for active participation? Often classroom conditions make it far too easy to choose non-participation (“opting out”). Although student participation is a graded component in many class syllabi, it is generally difficult for teachers to remember exactly who said what in each class. A few noteworthy students might stand out, but the majority easily become a blur. For such reasons, Bean and Peterson (2002) suggests that many instructors grade “participation” impressionistically and also rather unreliably.

Typical classrooms exhibit too many features of *zero-sum games* (Levine, n.d.) and not enough characteristics of *non-zero sum games*. In a zero-sum game there are always a fixed number of “winners” and “losers” and this is an appropriate paradigm for normative testing. Those who fall significantly below the statistical mean automatically fail, while those placing

above it receive high marks. Grading schemes with normative elements may enhance competitiveness among some individuals, but it's good to remember how this also automatically creates zero-sum conditions in which the number of students who pass determines how many must fail. In most teaching contexts, criterion-reference grading (also referred to as absolute grading) is more appropriate: theoretically every student can be a "winner". The classroom "commodities" of knowledge and satisfaction certainly are not *zero-sum* resources. If classrooms are played as *cooperative games* and interactants assist each other, all players benefit and become "wealthier". All too often, however, classrooms resemble *competitive games* in which students vie against each other as well as their instructors.

Whereas ideal classrooms may be described as *perfect information games* (Rasmusen, 2001, p. 47-50) in which students know exactly how a teacher will award grades and what is necessary to obtain such rewards, all too often teachers themselves are not entirely clear about how grades will be determined. Even if instructors are clear about grading issues, how can they be sure that their students have fully understood the information presented about it? The possibility of miscommunication needs to be acknowledged and to make classroom systems effective, often key information needs to be recycled. This adds *robustness* (Gawith, 2003) to the system, but can also slow down learning if there is too much recycling. In most classroom games (and life itself) time is a limited resource.

Most classrooms should be regarded as *Bayesian game systems* (Nurmi, 2005) in which information about the interactants involved is incomplete and outcomes are uncertain. To compensate for the lack of certainty, players make probabilistic guesses about what is most likely to happen based on past scenarios they have experienced. This creates an expectancy effect, and when educational courses are closely aligned to participants' expectations, then at least metaphorically a sort of *Nash equilibrium* (Rasmusen, 2001, p. 296-298) ensues.

In short, due to a variety of factors typical classes rarely represent *best-shot games* in which all participants are putting maximal effort into learning. Most likely, interactants are holding back to some degree - playing it "safe" by participating the required extent, but seldom beyond expected levels. The Japanese proverb "*deru kuge ga ataru*" [the nail which sticks out gets struck] epitomizes a tendency to conform to acceptable performance levels, conserve energy expenditures, and avoid risk.

Incentive grading systems

From a game theory standpoint, a variable that can be fairly easily manipulated in classroom contexts is the grading system. A wide amount of educational research has been conducted on alternative grading schemes. One that may be suited to students who are not operating in best-shot mode is a *point incentive system*. The idea behind it is purely behavioristic: reward behaviors which are considered desirable and, in some cases, punish those which aren't. There are many variations of this system. Before mentioning two scenarios I have used, three scenarios from universities in the United States will be briefly outlined.

Scenario 1

One of the more elaborate unitary point grading systems I am aware of is for a sociology course at the University of Hawaii by James and Nahl (1981). Students are awarded points in nine categories during the course. To pass the course students must earn 1,500 points. To obtain an "A" over two thousand points are needed. The grading scheme works on an honor code: students keep track of their own points and in some cases also decide how well they fulfilled specific activities. An interesting aspects of James and Nahl's system is their thought-provoking student feedback forms. Students rate their response to each lesson and to their daily lives in terms of six fixed-response prompts and two open-ended questions. One drawback of this system is that it presumes honesty. Since high-stakes decisions are associated with grading, in many cases some sort of teacher check is in order.

Scenario 2

In addition to basing their grades on term papers and test performance, Bean and Peterson (2002) also devote a section of their grades to student class participation. They developed a detailed six-point scoring rubric for class participation that lets students know in precise language how their performance will be rated. One thing Bean and Peterson emphasize is the importance of communicating grading standards clearly to students and, if necessary, modeling desired behaviors so that those in class get a very clear idea of how to participate. They also mention some useful strategies for encouraging shy students and ways to work with students who dominate discussions. Unfortunately, they do not provide a lot of specific information on precisely how they grade their courses.

Scenario 3

An interesting analytic point grading system can be found in Prof. Jeff Adams' sociology course at St. Michael's College. In his incentive system, points can be earned in four categories: quizzes (40% of the grade), applied papers (40% of the grade), chapter review questions (10% of the grade), and attendance-participation (10% of the grade). Points earned in one category cannot be transferred to another. An especially well-crafted feature of this system is the way that papers are graded according to a 4-point rubric in four areas. Hence, an outstanding report in all respects is awarded 16 points. It no doubt requires a data spreadsheet such as Excel[®] to work out this grading system for an entire class, but the overall system is well-balanced. What do students think of this grading system? That's hard to say with precision, but now many professors at universities in North America, the British Isles, Australia, and New Zealand are rated by review sites such as RateMyProfessors.com. 45 reviewers rated this professor's classes, giving him at 4.0 out of 5. Though it's beyond the scope of this paper to say what that actually suggests, the day when teacher evaluations are broadcast on the World Wide Web for all to see is already here in many parts of the world.

Incentive Grading System 1

In the spring of 2006 I adopted an incentive grading system in which points were awarded for desirable behaviors. Four behaviors were targeted: attendance, homework, presentations, and test results. To obtain course credit, students were notified from the first class that they had to earn at least 400 points by the end of the semester. 90% of that required minimum could be earned through attendance. However, merely being physically present in the class would not generate enough credit. To pass a course, students also needed to do homework, and/or prepare presentations, and/or score high in the exams. All grades were determined according to a simple formula, with As going to those with over 500 points, Bs to those garnering 450-499 points, and Cs to those who earned 400-449 points. The actual grading system is summarized in Table 1.

To implement this system, I printed up small point cards which were handed out to students whenever a desired performance was completed. To reduce the likelihood of photocopying such point cards, I made each one unique, stamped, and watermarked.

How did this grading system impact students in the six pilot classrooms? Two different information sources might shed some light on that question: teacher observations, and anecdotal student comments. Each are briefly discussed.

(1) Teacher observations

Four common response patterns to the Incentive Grading System 1 were observed in the six classes in which this system was implemented. The majority of students “played the game” well enough to earn 400 points, but were not much interested in earning extra points. Once they were confident of passing, they saw little reason to expend surplus energy seeking higher grades. This group of students could be described as *borderline-resters*. They were thrifty about their energy expenditures and worked hard up to the point of obtaining passing

Table 1 The 2006 spring and autumn semester incentive grading system

behavior	per unit value	max. per semester value
(A) Attendance		
Come to Class on Time	30 points per class	360 points if consistent
Come to Class 2-20 Minutes Late	20 points per class	240 points if consistent
Come to Class 21-40 Minutes Late	10 points per class	120 points if consistent
(B) Homework		
Workbook Tasks	5 points per page	up to 150 points
Composition Tasks	1 point per sentence	up to 165 points
(C) Presentations		
Weekly Presentation Tasks	Up to 15 points per task	up to 150 points
Semester Final Tasks	Up to 30 points per task	0 - 60 points*
(D) Tests		
All tests	1 point per correct response	30 - 200 points*

* This variable differed from class to class.

credits, but hardly further. In most classes, about two-thirds of the students seemed to fall in this category.

Another considerably smaller group of students continued to participate actively long after reaching the minimum benchmark. Quite likely such students were aspiring for high grades - or perhaps they simply enjoyed the subject.

A third group of students chose "opting out" by ceasing to come to class once they realized that they were too far behind to earn the requisite points. Since the incentive point system in Table 1 strongly emphasized attendance, students who missed more than three classes or were chronically late for class gradually understood that their chances of earning 400 points were slim. Perhaps 10% - 15% of the students in most classes exhibited these behaviors.

An even smaller group of students continued to attend and sometimes even participate actively even after any realistic hope of earning 400 points had vanished. Their behavior did not fit any theoretical pattern that I could understand. Were they still hoping to get credit or perhaps simply indifferent to their grade results? Bauwens, Lubrano, and Richard (2003) provide this advice about how to regard such misfitting cases which counter theoretical expectations -

There now exists a large body of economic experiments that seem to suggest that economic agents (mostly students) when confronted with experimental scenarios often behave in ways that appear to contradict theory prescriptions. Whether or not such experimental evidence actually contradicts theory is often a matter of interpreting the agent's perceptions of the situation they are confronted with; this comment applied a *fortiori* to real life situations . . . (p. 5)

(2) Student comments

Quite a few students complained about having to keep track of so many point cards. By the end of the semester, some had accumulated well over fifty cards worth over six hundred points. At least one student in nearly every class lost their cards. To minimize such problems, by mid-term I allowed students to turn in their cards at the end of each class for recording. Though this was time-consuming, worries about card loss were reduced.

The most common comment about the Incentive Grading System 1 was that it was overly strict about attendance; many students seemed to prefer much looser attendance standards. Only a few actually praised the Incentive Grading System 1 for being "fair" or "consistent".

Incentive Grading System 2

By spring semester-end I sensed a need to simplify the grading system. In the autumn of 2006 a new system was adopted which varied from the previous one in three ways.

First, instead of awarding individual printed point cards, students wrote how many incentive points they earned on a special mark sheet which appears in Appendix A. At the

end of each class, this mark sheet was collected and I checked student marks to make sure they were correct. This eliminated the possibility of point card loss and also gave students more precise feedback about how they stood in terms of fulfilling credit requirements.

Another new feature of the Incentive Grading System 2 is that more time for cognitive processing and feedback was allocated at the end of each class. In addition to recording how many points they earned, students also wrote down the new vocabulary that they encountered and rated one different aspect of the lesson each time. This made it possible to get a window into not only what students were learning, but also how they were feeling.

The third difference was a shift in the way points were awarded: I began to focus on the *quality* of the spoken responses and homework instead of merely the *quantity*. Pragmatically rich responses earned more than ones that were shallow.

How did students respond to this revised system? Again, so far only teacher observations and anecdotal student comments offer a glimpse. The observed behaviors were essentially the same as those found in the prior grading system, but the response ratios shifted: more students became either *high-achievers* or *borderline-resters* than before. Though more than a few students still complained about the way this system was too strict about attendance, in general there were fewer complaints.

Is it fair to say that Grading System 2 represents an improvement on the previous system? In the context of this limited case study, the evidence suggests so. However, in other teaching contexts it might be an altogether unnecessary contrivance. By defining standards for student performance so clearly, it perhaps discounts the value of interactions which are not graded. Some students, for example, probably learn more from pre-class or after-class interactions with the teacher and each other than they do from the lesson. Such interactions are not within any grading scheme. By defining what types of performance should be graded too narrowly, this assessment scheme fails to measure important facets of domain knowledge. In short, the constant process of awarding points for interactions has a tendency to trivialize the interactions.

Discussion and Conclusion

Before adopting any assessment system, teachers need to have a clear rationale for why they grading in the first place. The same standards which Bachman and Palmer (1996, pp. 17-18) have suggested for testing in general also apply to grading in particular. The Grading System 2 mentioned in this paper could justly be criticized in terms of its lack of authenticity and construct validity, though scrapping it entirely brings us back to the scenario of the “typical classroom” mentioned earlier. Both grading schemes mentioned here are heuristic devices that still need refinement. As Frisbie and Waltman (1992) suggest, teachers’ approaches to grading often evolve over time as their ideas about teaching itself change in response to situations they encounter. Though the grading models described here are probably not needed for students keenly motivated to master a target domain, they might provide an instru-

mental incentive for those who aren't. However, reflective teachers also need to look deeper and consider what factors foster non-responsiveness in classes and how those can be reduced. Though grading might be part of the formula, it is by no means the only part - or possibly even the most important. This particular study has an structural modeling bias, yet it should not be denied that other important features such as teacher rapport, personal goal-orientation, and peer acceptance also powerfully mold behavior.

This paper has outlined two incentive point grading systems. Further studies about this issue need to address the following questions: (1) If a point grading system is adopted, how should the cut-off points be rationally set?, and (2) Would an analytic scale be superior to a unitary one? A brief comment on each point is in order.

(1) Determining the cut off points

How lenient or stringent should the standards for passing a university level class be? As Japan's population shrinks and many universities soften their admission standards, this question becomes increasingly relevant. Allowing students to obtain up to 360 of the required 400 points needed to pass a course is undoubtedly lenient. As a consequence, even those who understand hardly any English can still make it through a course by attending regularly, doing a bit of homework, and using effective guessing strategies during tests. Only a few universities in Japan seem to link grades foreign language course to clear "can-do" benchmark performance standards or widely-known external tests. The Tokuyama College of Technology, for example, requires students to obtain a minimum TOEIC[®] score of 400 to graduate. Although the Japanese Ministry of Education, Culture, Sports, Science, and Technology's 2003 Action Plan encourages the use of external proficiency examinations in screening those *entering* high schools and universities, the performance criteria for *completing* many courses remains fuzzy.

Would raising the performance bar from 400 points to, say, 432 points promote higher standards? Or would it merely foster some sophisticated game playing in which students go through more motions to obtain points without actually learning anything? Subsequent research is needed to answer that question.

Where should the benchmarks for other grades be set? Both systems described in this paper had rather arbitrary cut-off points for each grade level. Clearer benchmarks for each performance level should be set and these should be translated in rubrics that students understand. As Stiggins (1997) emphasizes, the focus in class should not be exclusively on grading, but on communication. And such communication should not be something that occurs merely at the end of a semester: some type of *micro-feedback* should be given to interactants each lesson.

(2) Unitary vs. analytic scaling

Both of the incentive grading systems described in this paper used unitary point

scales. Such systems have an advantage of simplicity. Moreover, students who are weak in terms of one facet can compensate by making doing more in another. However, it might be worth adopting a hybrid analytic scale which requires students to earn a certain number of points from tests and manifest certain tokens of effort in specified categories. In such a scenario "attendance points" would be distinct from "presentation points" or "test points". However, until students actually see the reason for the target behaviors and internalize the need for activities such as speaking up, attending regularly, or doing homework any amount of "system engineering" will likely have limited long-term effects. Proficient learners tend to do many of the activities which are rewarded in incentive point systems automatically. Trying to "jump start" ordinary students into adopting those same behaviors takes considerable prodding. Perhaps that is why educators such as Carrell (1998) repeatedly emphasize the need for meta-cognitive training as a way for learners to explore their own learning processes.

References

- Adams, J. (2006). *How you will be graded in this course*. Accessed Dec 12, 2006 at <http://academics.smcvt.edu/jadams/social/Social%20Grading%20System.htm>.
- Bachman, L. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baker, R. S., et al. (2005) *Do performance goals lead students to game the system?* Accessed Dec 1, 2006 at <http://www.psychology.nottingham.ac.uk/staff/lpzrsb/BRCKAIED2005Final.pdf>.
- Bauwens, L., Lubrano, M., & Richard, J. F. (2003). *Bayesian inference in dynamic econometric models*. Oxford / New York: Oxford University Press.
- Bean, J. C., Peterson, D. (2002, Oct. 19). *Grading class participation*. Accessed December 1, 2006 at http://academicaffairs.csufresno.edu/assocprovost/documents/pdf/grading_class_participation.pdf.
- Carrell, P. L. (1998, March). Can reading strategies be successfully taught? *The Language Teacher*, 22 (3) 1-9. Accessed Dec 15, 2006 at <http://jalt-publications.org/tlt/files/98/mar/carrell.html>.
- Covaleskie, J. F. (1994, February 10). The educational system and resistance to reform: The limits of policy. *Education Policy Analysis Archives*, 2 (4). Accessed Dec 13, 2006 at <http://epaa.asu.edu/epaa/v2n4.html>
- Foundation for Teaching Economics. (2006). *Teacher background to game theory and experimental economics*. Accessed Dec 11, 2006 at <http://www.fte.org/capitalism/activities/ultimatum/appendix1/>.
- Frisbie, D. A., Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practice*, 6 (4) 29-37. Accessed Dec 14, 2006 at <http://depts.washington.edu/grading/plan/frisbie1.htm>.
- Gawith, J. (2003). *The fuzzy front end: A key step to successful projects*. Accessed on Nov 8, 2006 at <http://www.techednz.org.nz/proceedings/2003/Proceedings110-149.pdf>.
- James, L. & Nahl, D. (1981). *Psychology* 222. Accessed on Sep 10, 2006 at <http://www.soc.hawaii.edu/LEONJ/LEONJ/leonpsy/instructor/society/society.html#5>
- Japanese Ministry of Education, Culture, Sports, Science, and Technology (2003). *Regarding the establishment of an action plan to cultivate "Japanese with English Abilities."* Accessed Oct 3, 2006 at <http://www.mext.go.jp/english/topics/03072801.htm>.

- Levine, D. K. (n.d.). *What is a solution to a zero-sum game?* Accessed Dec 8, 2006 at <http://levine.ssc-net.ucla.edu/Games/zerosum.htm>.
- Nurmi, P. (2005). Bayesian game theory in practice: A framework for online reputation systems. *University of Helsinki Department of Computer Science Series Publications C Report C-2005-10*. Accessed Dec 13, 2006 at http://www.cs.helsinki.fi/u/ptnurmi/papers/nurmi_bayesian_games_reputation.pdf.
- Osborne, M. J. (2004). *An introduction to game theory*. Oxford: Oxford University Press.
- Rasmusen, E. (2001). *Games & information*. (3rd Ed.). Malden, Oxford, Victoria: Blackwell Publishing.
- RateMyProfessors.com. (2006) St. Michael's College Ratings Page. Accessed Dec 12, 2006 at <http://www.ratemyprofessors.com/SelectTeacher.jsp?sid=860>.
- Stiggins, R. (1997). *Student-centered classroom assessment*. (2nd ed.) Columbus OH: Merrill, an imprint of Prentice Hall.
- Tokuyama Senmon Gakkou. (2006). *College Bulletin 2006*. Accessed Dec 11, 2006 at <http://www.tokuyama.ac.jp/japanese/information/introduction/images/catalogue2006.pdf>.

Appendix A. Sample student mark sheet used for the Autumn 2006 semester.

NOTE: This version is abbreviated. The full version is online at <http://www.tnewfields.info/Articles/ingradApA.htm>

Student Mark Sheet

NOTE: Hand this back to the teacher at the end of each class.

Name: _____ Student #: _____

Your Committee (circle one): Review / International News / Domestic News /
TOEIC® Listening / TOEIC® Reading / Feedback-Social

My goal for this class (circle one): ___ 400-449 points (Grade C) ___ 450-499 points (Grade B)
___ 500+ points (Grade A) ___ The Top Score (Grade S)

Lesson 1. (10/5) Attendance: _____ Participation: _____ Today's Total: _____

How easy was today? too easy -2 -1 0 -1 -2 too difficult [0= just right]
New words I encountered: _____

Lesson 2. (10/12) Attendance: _____ Participation: _____ Homework: _____
Today's Total: _____ Sub-Total So Far: _____

How interesting was today? (circle one) very boring -2 -1 0 +1 +2 very interesting
New words I encountered: _____

Lesson 3. (10/19) Attendance: _____ Participation: _____ Homework: _____
Today's Total: _____ Sub-Total So Far: _____

The teacher's speech is: (circle one) too fast -2 -1 0 +1 +2 too slow [0= just right]
New words I encountered: _____

Lesson 11. (1/11) Attendance: _____ Participation: _____ Homework: _____
Today's Total: _____ Sub-Total So Far: _____

Where do you rank yourself in this class? (circle one) Near the Bottom - In the Middle - At the Top
New words I encountered: _____

Lesson 12. (1/19) Attendance: _____ Participation: _____ Homework: _____
Today's Total: _____ Sub-Total So Far: _____

How helpful is this course? (circle one) not helpful -2 -1 0 +1 +2 very helpful
New words I encountered: _____

Final Test: _____ points

TOTAL POINTS: _____ RANK: _____ GRADE: _____

REMINDER: Be sure to hand this back to the teacher at the end of each class.