An Analysis of Negative Marking in Multiple-Choice Assessment

Dr Alan Holt
The School of Computing and Mathematical Sciences
The University of Waikato, Hamilton, New Zealand

aholt@cs.waikato.ac.nz

Abstract

This paper examines guessing strategies in multiple-choice assessments. It first looks at the likelihood of achieving a pass score given that the student has no knowledge of the subject area and adopts a *pure* guessing strategy. Then an analysis of *negative marking* schemes, as a means of discouraging guessing, is presented and used in order to examine the trade-offs between taking a guess and not answering a question.

The analysis shows that with a negative marking scheme the benefits/detriments of guessing depend upon the severity of the penalty of an incorrect answer relative to level of reward for a correct answer *and* the number options from which the students have to choose. From this analysis it is possible to calibrate the marking scheme and establish "fair" penalties.

Keywords: Multiple-choice assessment, negative marking, guessing strategies.

1 Introduction

The analysis in this paper was prompted by a course taught by the author at a former academic institution. The course relied heavily on multiple-choice assessments, where *both* negative and non-negative marking schemes were adopted. That is, for some tests, wrong answers were penalized with a negative mark while in other tests wrong answers were awarded zero (in both cases unanswered questions were given zero).

The course in question, Computer Systems Organisation, was a compulsory first year Computer Science undergraduate course. Two out of the four course assessments were multiple-choice tests and the end of year examination was a combination of essay questions *and* multiple-choice questions.

The rationale behind using a mixture of marking strategies was never fully understood by the author, but it appeared that negative marking was used during the early stages of the course as a *shock and awe* tactic (Havenar 2003) to "encourage" students to study. While, in the latter stages of the course, negative marking was dropped to give struggling students a chance.

According to Nicholls, (2002) multiple-choice assessments offer a number of advantages:

- large areas of the syllabus can be covered
- many cognitive abilities can be tested
- marking can be carried out rapidly by someone with no knowledge of subject or by machine
- careful analysis of the test results is possible
- individual item difficulties and discriminations can be calculated
- high reliability is possible
- speedy reporting of assessment results is possible

The third bullet item in the list above is particularly significant. Being a compulsory first year course, student numbers were very high (approximately 200), thus automated marking was very attractive and a strong motivator (for the lecturing staff) for adopting multiple-choice assessment. Nicholls (2002) also outlines the disadvantages of multiple-choice testing:

- good items are difficult to write and not everyone is capable of writing them
- a lot of time is required to construct a test
- guessing is encouraged, but this can be corrected for

With regard to last bullet item, section 3 of this paper presents an analysis of guessing in order to gain some insights into the likelihood that a student with little or no aptitude for the subject area can completely bypass the learning objectives and pass the test by purely guessing. Section 4 presents a formal description of a negative marking scheme which is used in Section 5 to carry out an analysis of varying degrees of "aggressiveness" in guessing strategies with respect to a student's expected grade. In Section 6 it is shown how (negative) marking schemes can be calibrated in to order set "fair" penalties.

This paper focuses on negative based marking of multiple-choice assessment because that is the scheme that was used for the Computer Systems Organisation course. However, it is acknowledged that there are other schemes for discouraging guessing. For example, confidence-based schemes (Gardener-Medwin and Gahan 2003) require students to provide a measure of certainty that an answer is correct, which is taken into account in when awarding a mark. Each marking scheme will have its own issues and is subject to the type analysis applied to negative-based marking presented in this paper.

This quality assured paper appeared at the *19th Annual Conference of the National Advisory Committee on Computing Qualifications* (NACCQ 2006), Wellington, New Zealand. Samuel Mann and Noel Bridgeman (Eds). Reproduction for academic, not-for profit purposes permitted provided this text is included. www.naccq.ac.nz

2 Terminology

Each question on the multiple-choice test is called a *stem*. Associated with each stem are a number of *options*, which are a set of alternative answers from which the student must select. These options consist of a *key* (the correct answer) and a number of *distracters* (incorrect answers). Some multiple-choice assessments have stems, which have multiple keys. Such assessments are not considered, only assessments with single keys are analysed in this paper

3 Analysis of Guessing

Given that there are a finite number of options, θ per stem, the probability of guessing the answer correctly (selecting the key) is $p = 1/\theta$. This is based upon the assumption that all options are equally believable. However, equally believable distracters may be difficult to achieve if θ is high. Indeed Hayadyna and Downing (1993) argue that there are only one or two effective distracters in approximately in approximately two thirds of multiple-choice questions.

Furthermore, while a student may not be entirely confident of the correct option, he or she may know that a proportion of θ are incorrect, either because they have some knowledge of the subject area or some of the distracters are not believable (or both). This effectively reduces the available options and increases the chances of selecting the correct answer by "chance", that is $1/\theta < p \leq 1$. Nevertheless, for simplicity $p = 1/\theta$ is assumed in this analysis.

We consider only multiple-choice tests with one key per stem. A student with no knowledge of the subject area, can, by adopting a purely guessing strategy achieve a pass grade G (of say) $i = 0.4$ (that is 40%) with a probability $P(G \geq i)$, where $P(G \geq i)$ can be derived from a Binomial distribution function (Staek and Woods 1994):

$$P(G \geq i) = \sum_{k=i}^{k=n} P(G = k) = \sum_{k=i}^{k=n} \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{(n-k)} \quad (1)$$

and n is the number of stems in the test. Figure 1 shows the probability of a student achieving a pass grade $P(G \geq i)$ (for $i = 0.4$) using a purely guessing strategy. This assumes that 1 mark is awarded for selecting a key and a zero mark for selecting a distracter. As the student is adopting a guessing strategy it is assumed that all questions are attempted. The surface plot in Figure 1 shows the level of $P(G \geq i)$ over a range of $n = 10,...,100$ (the number of stems in the test) and a range of options for each stem $\theta = 2,...,10$.

It can be seen that for a low number of stems and options the probability of passing is quite high. However the probability of a pass diminishes exponentially as the number of stems and options increases. In order to discourage "passing by guessing" (or rather passing by luck) the solution would appear to be, to have a many stems and many options. There is, however a logistical problem with this solution, just as the probability of passing decreases exponentially with the number of stems and options, so too does the increase in the amount of work required to compile the test (outlined in the second item of the "disadvantages" list in Section 1 (Nichols 2002)).

4 Negative Marking Schemes

Realistically, not many students would adopt a pure guessing strategy. Rather they would take a multiple-choice assessment with a certain amount of understanding of the subject area and use guessing as a means of enhancing their grade.

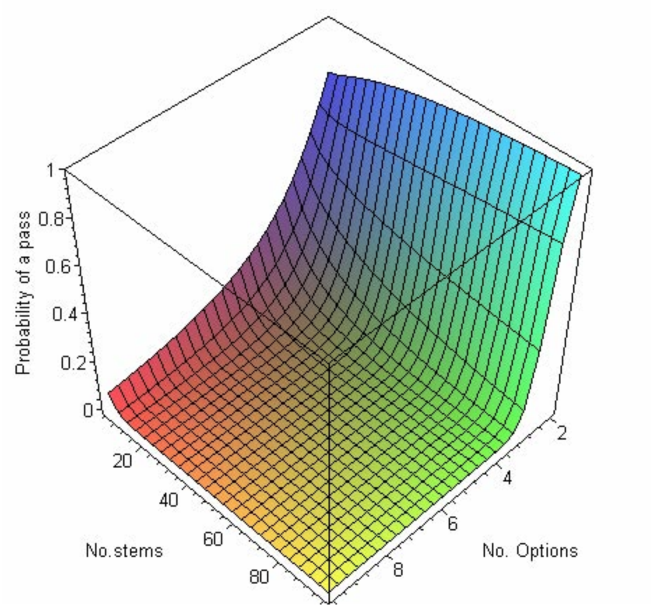

Figure 1 The probability of passing using a pure guessing strategy

Negative marking schemes provide a method of discouraging students from guessing. Negative marking schemes reward key selection (correct answers) but penalize distracter selection (incorrect answers). Thus a negative marking scheme would award a mark M according to:

$$M = \begin{cases} a & \text{key selection} \\ -b & \text{distracter selection} \\ 0 & \text{stem unaswered} \end{cases} \quad (2)$$

where $a > 0$ is the level of the "reward" and $b \leq 0$ is the level of the penalty. Unanswered stems are awarded a mark of zero. Although this is a model of a negative marking scheme, the scheme is generic enough to encompass a no-penalty marking scheme, as b can be zero.

On completing the test a student will have answered a proportion ρ of the stems correctly, and a proportion $1-\rho-U$ of the stems incorrectly, where U is the proportion of the stems that the student did not answer. The normalized grade $b/a \leq G \leq 1$ is given by the expression below:

$$G = \frac{a\rho + b(1-\rho-U)}{a} \quad (3)$$

Note that if $b<0$ it is possible for the student to achieve a negative mark (which would detract from the aggregate mark for other assessments in that unit).

5 Analysis of Negative Marking

From the analysis above, let us assume that a student is capable of deducing the correct answers for a proportion ρ of the stems and leaves U unanswered. The remaining $1-\rho-U$ stems are attempted but the student is uncertain of the answer. However only in the worst-case scenario would the student select distracters for all $1-\rho-U$ of these stems. On average a student should select $\theta^{-1}(1-\rho-U)$ proportion of keys by random chance. Similarly a student will select a proportion of distracters: $(1-\theta^{-1})(1-\rho-U)$.

Putting this together with the marking scheme in Section 4 a student's *expected* (normalized) grade is given by the expression below:

$$E[G]=\frac{a\rho+a\theta^{-1}(1-\rho-U)+b(1-\theta^{-1})(1-\rho-U)}{a} \quad (4)$$

Figure 2 shows three surface curves of the expected student grade for the following proportions of unanswered stems $U=0$, $U=0.2$ and $U=1-\rho$. $U=0$ represents an "aggressive" guessing strategy. That is, key selection for all $1-\rho$ stems, which lie outside the student's understanding of the subject area, are guessed. $U=0.2$ represents a more conservative guessing strategy, while for $U=1-\rho$, no key selection is left to random chance. For the purpose of the analysis the proportion of stems the student can correctly answer is set to $\rho=0.55$ and the number options is set to $\theta=5$.

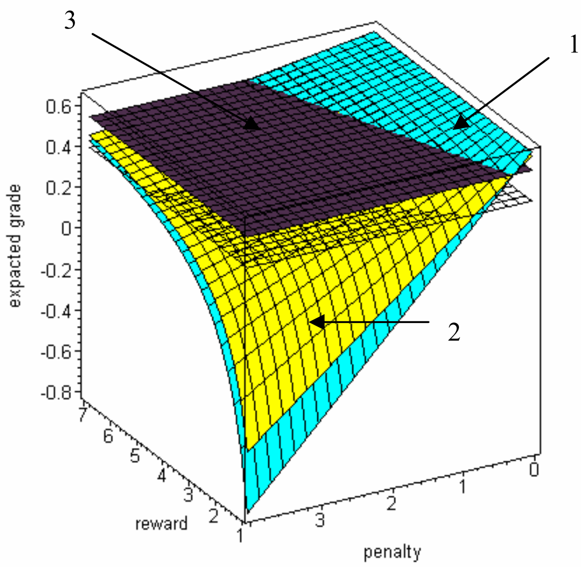

Figure 2 Expected student grade: guessing and negative marking

Surface 3 in Figure 2 is flat (unvarying) across the range of rewards ($1\le a\le 7$) and penalties ($0\le b\le 5$). This is because only the stems for which the student is confident of selecting the key are attempted (that is $U=1-\rho$). While not penalized for guessing, the student does not take advantage of key selection by random chance.

The advantages of guessing are evident from surfaces 1 and 2. Students that guess when the penalties are low (or zero) supplement their expected grades (inversely proportional to U) relative to students that do not guess. However, as penalties increase relative to rewards the expected grades diminish dramatically, beyond the 0.4 pass grade (denoted by the wire frame mess) and even into negative scores. It is clear that low penalties favour aggressive guessers over conservative guessers. However, the converse is true as penalties increase.

6 Fair Penalties

In this section *fair* penalties for discouraging guessing in negative marking schemes are examined. Ideally, we want to reward ability rather than luck. Nevertheless excessive penalties for incorrect answers could intimidate students. Recall that, with a negative marking scheme, a negative score will detract from scores on other assessments. Consider the two extreme cases of an *aggressive* guesser ($U=0$) and a non-guesser ($U=1-\rho$). By substituting $U=0$ into equation (4) an expression for the expected grade $E[G_{U=0}]$ of an *aggressive* guesser can be found:

$$E[G_{U=0}]=\frac{a\rho+a\theta^{-1}(1-\rho)-b(1-\theta^{-1})(1-\rho)}{a} \quad (5)$$

Likewise by substituting $U=1-\rho$ into (4) the expression expected grade $E[G_{U=1-\rho}]$ for the non-guesser reduces to:

$$E[G_{U=1-\rho}]=\rho \quad (6)$$

If we consider a fair penalty b_{FAIR} to be that which would have neither beneficial nor detrimental effects on a student's grade (that is when $E[G_{U=0}]=E[G_{U=1-\rho}]$), then the solution to the equation:

$$\frac{a\rho+a\theta^{-1}(1-\rho)-b_{FAIR}(1-\theta^{-1})(1-\rho)}{a}=\rho \quad (7)$$

yields the expression:

$$b_{FAIR}=\frac{a}{\theta-1} \quad (8)$$

So for the example above where $\theta=5$, the penalty would have to be one quarter of the reward. More formally $b_{FAIR}=a/4$. The graph in Figure 4 shows the fair penalty for a range of options $2\le\theta\le 20$ and a range or rewards $1\le a\le 10$. It can be seen that for a small of number options, the penalties have to be *stiff*, that is numerically equal to the rewards. However, for a large number of options penalties are negligible which is consistent with the results of the "pure guessing strategy" analysis presented in Section 2. For the Computer Systems Organization course, the number of stems was $\theta = 5$, and a reward/penalty scheme of $\{a = 3, b = 1\}$ was adopted. In view of the results of the analysis above this would appear to be a somewhat harsh marking scheme.

7 Conclusions and Future Work

Negative-based marking is by no means the only scheme for discouraging guessing, but each scheme introduces its own issues and could expose "secondary" abilities (or inabilities) in students. By secondary abilities we mean those not directly associated with course subject. For example, a negative marking scheme with low penalties may favour aggressive guessers or a confidence-based scheme, depending upon how it is implemented, may penalise those who are honest about their self-belief. We may value intelligent guessing in property speculators, but not in designers of safety critical systems.

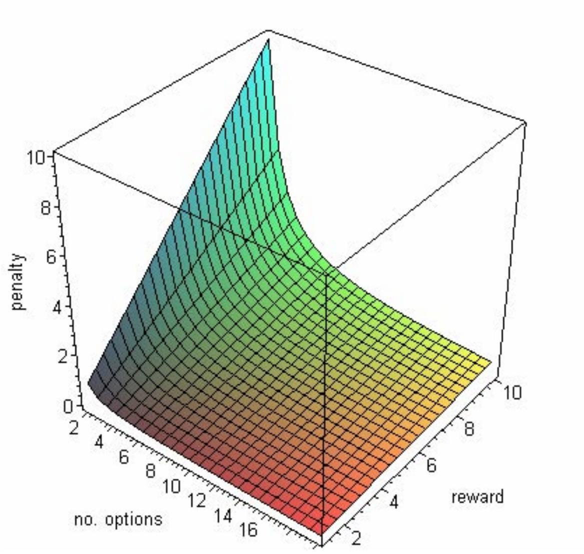

Figure 3 Setting "fair" penalties

While the focus of the analysis presented in this paper has been on negative-based marking scheme, we feel confident that these models could be adapted for other schemes and could be used to gain some insight into their effectiveness.

Multiple-choice assessments are open to abuse through guessing. With essay based questions, a student with little or no understanding of the subject area that resorts to guessing only has a "shot in the dark", while with multiple choice assessment the odds, though long, are not negligible.

Multiple-choice assessments with a high number of stems and options reduce the chances of a student of attaining a pass grade by purely guessing. However, this increases the amount of work involved in constructing a multiple-choice test. One also has to consider, if the number of stems is high, the time allocation for taking the test must also be high.

Negative marking schemes are designed to discourage guessing. Correct answers are rewarded, wrong answers are penalised while unanswered questions are neither (awarded zero). It has been shown that *fair* penalties are related to then number of stems and the level or reward for a correct answer, *but* not on the number of options. An expression has been derived such that penalties for incorrect answers can be computed that are (on average) neither beneficial nor detrimental to a student's overall grade.

8 Acknowledgements

The author wishes to thank Dr Adrian Davies and Dr Chi-Yu Huang for their many helpful comments and suggestions.

9 References

Bloom, B.S. (ed.) (1956): "Taxonomy of educational objectives". Handbook 1. The Cognitive Domain. New York, David McKay.

Gardener-Medwin A.R. and Gahan M.,(2003) "Formative and Summative Confidence-Based Assessment". Proc. 7th International Computer-Aided Assessment Conference, UK, pp. 147-155, July 2003

Haladyna, T. and Downing, S. (1993) "How many options is enough for a multiple choice test item?" Educational & Psychological Measurement journal Vol 53, No. 4, pp. 999-1010, 1993.

Havener J (2003) "Shock and Awe" Battle Plan for Iraq, (http://www.uglatto.com/files/Shock%20and%20Awe.pdf

Nicholls, G. (2002) "Developing Teaching and Learning in Higher Education", *Routledge Farmer*, 2002,

McFarlane A. (1997) "Information Technology and Authentic Learning", Routledge, 1997.

Staek H. and Woods J.W., (1994) "Probability, Random Process, and Estimation Theory of Engineers", Prentice Hall, Second Edition.

Sparkes J.J. (1995) "Quality in higher education, IEE Engineering Science and Eduaction", Journal, Vol. 4, No. 6, pp, 253-258, December 1995.

Tanenbaum A.S. (1999) "Structured Computer Organization", Prentice Hall. Fourth Edition.

Woodford K. and Bancroft P. (2005) "Multiple Choice Questions not Considered Harmful", Australian Computing Education Conference, Vol. 42, 2005.