



Scoring methods for multiple choice assessment in higher education – Is it still a matter of number right scoring or negative marking?



Ellen Lesage^{*}, Martin Valcke¹, Elien Sabbe²

Zomerstraat 60, 9000 Ghent, Belgium

ARTICLE INFO

Article history:

Received 31 October 2012

Received in revised form 7 July 2013

Accepted 8 July 2013

Keywords:

Assessment
Higher education
Multiple choice

ABSTRACT

In higher education, a multiple choice test is a widely known format for measuring student's knowledge. The debate about the two most commonly used scoring methods for multiple choice assessment – number right scoring (NR) and negative marking (NM) – seems to be a never-ending story. Both NR scoring as NM do not seem to meet the expectations. However, available research hardly offers alternative methods. Clearly, there is a growing need to explore these alternative scoring methods in order to inform and support test designers. This review aims to present an overview of (alternative) scoring methods for multiple choice tests, in which strengths and weaknesses of each method are provided.

© 2013 Elsevier Ltd. All rights reserved.

Contents

Purpose of this review	188
Methodology	189
Conventional scoring methods	189
Number right scoring	189
Negative marking	189
A preliminary conclusion about number right versus negative marking	190
Non-conventional scoring methods	190
Scoring methods that acknowledge partial mastery	190
Multiple answers	190
Retrospective correcting for guessing	191
Standard setting	191
Conclusions	191
References	192

Purpose of this review

The major goal of testing is to extract student's knowledge of the subject matter from responses to tests. The test score has to correspond as close as possible to the 'true' level of knowledge mastery of students. The debate about scoring multiple choice test formats presents a particular case in the research literature. Many

studies have compared scoring methods, however, the comparison involved tests from varying knowledge domains, neglected differences in test length and/or applied different analysis criteria. As a consequence, comparisons of multiple choice test approaches turned out to be difficult and presented mixed results. In addition, the research conclusions neglected the particular impact of internal and external variables. The specific educational context is a key external variable. Internal variables are linked to student variables such as risk-taking behavior and test anxiety. Not surprisingly, an unequivocal, evidence-based advice concerning the scoring of multiple-choice tests does yet not exist.

In the present paper, we do not continue the debate whether the two most commonly scoring methods, number right (NR) or negative marking (NM), is to be preferred. We rather aim at

^{*} Corresponding author. Tel.: +32 477455752.

E-mail addresses: ellenlesage@hotmail.com (E. Lesage),

Martin.Valcke@UGent.be (M. Valcke), Elien.Sabbe@UGent.be (E. Sabbe).

¹ Tel.: +32 92648675.

² Tel.: +32 93310062.

presenting a comprehensive review of multiple choice scoring methods currently used in higher education, resulting in an analysis of weaknesses and strengths. Appendix 1 summarizes the multiple choice scoring methods, commonly used in previous studies from 1979 onwards. We compare these scoring methods along several dimensions that are important to multiple choice tests: scoring rule, ideal answer strategy, reliability and validity, students' guessing behavior, partial knowledge and educational issues.

Methodology

The online databases of ISI Web of Science, the Educational Resources Information Center (ERIC), Science Direct and Google Scholar were searched between the years 1979 and 2011. The following keywords were used: *multiple choice; assessment; examination; higher education; scoring method; number right scoring; negative marking; formula scoring; correction for guessing; standard setting; guessing or risk-taking behavior*. All abstracts were reviewed and relevant documents were selected. These relevant articles were reviewed and their references were searched for associated articles.

In the context of the review of the research, in particular the knowledge domains of the medical sciences was covered, and this in relation to the topics standard setting and scoring methods that acknowledge partial mastery. In addition, specially the knowledge domains psychology, economics, law and mathematics appeared in the review articles.

Conventional scoring methods

Number right scoring

Traditionally, multiple choice tests have been scored using a conventional number right (NR) scoring method (Bereby-Meyer, Meyer, & Flascher, 2002; Kurz, 1999). Correct answers are scored with a positive value, incorrect answers and absent or omitted answers with a value of zero. The sum of the scores for correct responses is the test score. A major concern about this scoring method, is that students can answer correctly through guessing (Choppin, 1988). Students without the ability to solve a particular item, gain marks by guessing (Budescu & Bar-Hillel, 1993; Frary, 1988; Kubinger, Holoher-Ertl, Reif, Hohensinn, & Frebort, 2010). Guessing introduces a random factor into test scores that lowers reliability and validity (Bereby-Meyer et al., 2002; Burton, 2001; Kubinger et al., 2010; Prihoda, Pinckard, McMahan, & Jones, 2006). Test designers can as such not distinguish between correct answers based on knowledge mastery versus those based on a guess (Bar-Hillel, Budescu, & Attali, 2005).

Negative marking

Various scoring formulas have been presented to correct for guessing (Kurz, 1999). The 'rights minus wrongs' correcting model (Kurz, 1999) is predominant and penalizes the student for incorrect responses. The fundamental idea behind this scoring method, is that students acknowledge they will lose marks for incorrect answers (Betts, Elder, Hartley, & Trueman, 2009). As a consequence, students are discouraged to guess, and this is expected to increase test reliability and validity because the test score is a truer reflection of a student's ability (Kurz, 1999). Comparable to NR, a correct response results in a positive score and omitted items result in no mark (Hammond, McIndoe, Sansome, & Spargo, 1998). A straightforward interpretation is that the expected total score should be zero if a student guesses all answers at random. For this to happen, the penalty for an incorrect

answer should be $1/(n - 1)$, where n stands for the number of choices (Karandikar, 2010). We use the term 'negative marking' to describe this scoring method.

An alternative other model, proposed by Traub, Hambleton, and Singh (1969), rewards a student for not guessing by awarding points for omitted items rather than penalizing for incorrect responses. This presents a psychological advantage since it rewards the desired behavior rather than penalizing undesirable behavior (Crocker & Algina, 1986). Prieto and Delgado (1999) favor the latter scoring rule as the best of the scoring methods discussed thus far, and refer to resulting performance indicators and score reliability. Students do not feel threatened by receiving a reward for skipping items, as compared to receiving a penalty for incorrect responses.

Studies (Burton, 2002; Muijtjens, van Mameren, Hoogenboom, Evers, & van der Vleuten, 1999) report an increase in validity or reliability when negative marking is implemented. However, these studies only show slight improvements (Kurz, 1999) and they specifically examine true/false/items. Because of the high susceptibility to guessing, the results cannot simply apply to items with more alternative choices. Moreover, there is no consistency concerning applied analysis criteria, test length, knowledge domains and test instructions in these studies.

There remain other concerns about the use of negative marking. The rationale behind this scoring method – discouraging guessing behavior – seems to miss its purpose. It does not solve the guessing problem (Bar-Hillel et al., 2005; Betts et al., 2009); it even introduces new problems that are not observed in the case of NR scoring. Students differ in their guessing behavior: some students dare to take more risks than others. This introduces a first concern about students' risk attitudes adding to uncontrolled sources of variance, thus reducing the test's reliability and validity (Bar-Hillel et al., 2005). Moreover, guessing might benefit students who guess frequently compared to students with equal ability levels who do not guess (Choppin, 1988; Muijtjens et al., 1999). Some authors argue that by implementing negative marking, multiple choice tests rather measure students' answering strategies and risk-taking behavior instead of the mastery of domain knowledge (Budescu & Bar-Hillel, 1993; Choppin, 1988; Fowell & Jolly, 2000; Hammond et al., 1998; Kurz, 1999; Moss, 2001; Prihoda et al., 2006).

A third concern is related to the instruction to be given to students about guessing behavior. When negative marking was first introduced, students were simply advised not to guess (Davis, 1967; Frary, 1988). Change came when students were instructed to guess, whenever they could eliminate at least one or more alternative choices (Betts et al., 2009; Davis, 1967; Frary, 1988; Hammond et al., 1998). The study of Betts et al. (2009) suggests that guessing does not significantly reduce student performance. Hammond et al. (1998) advise instructors to be very cautious about instructing students not to guess. However, Budescu and Bar-Hillel (1993) point out the contradictory character of these instructions to guess, since the underlying principle of this formula is to discourage guessing. These authors conclude it is almost impossible to give recommendations that are fair and beneficial to all students. Students react in inconsistent ways, thus the question whether we have to instruct a student to guess or not is therefore far more difficult to answer than it seems (Budescu & Bar-Hillel, 1993). In this respect, it is also difficult for students to figure out the optimal decision strategy under negative marking.

Finally, there is confusion about the amount of the penalty given for incorrect answers (Karandikar, 2010). Some authors (Budescu & Bar-Hillel, 1993; Espinosa & Gardeazabal, 2010) state that an effective penalty that effectively discourages guessing, should exceed the standard penalty of $1/(n - 1)$.

Download English Version:

<https://daneshyari.com/en/article/372669>

Download Persian Version:

<https://daneshyari.com/article/372669>

[Daneshyari.com](https://daneshyari.com)