Selected response assessment formats versus constructed response assessment formats in statistics education. Pragmatic and paradigmatic considerations in choosing a type of assessment

Gilberte Schuyten
University of Gent, Department of Data Analysis
Henri Dunantlaan 1
B-9000 Gent, Belgium
Gilberte.schuyten@rug.ac.be

1. Introduction

In the 1990s the term "educational assessment" replaced the term "educational testing" (Hambleton, 1996, p.899). Student assessment shifts in emphasis from to make educational decisions such as passing score for graduation towards to give feedback, from assessing basic skills towards performance skills and from objective assessment towards performance assessment. The answer given on this why, what and how question shows a rich variety of purposes, of skills and of formats. Webb (1992) discusses two particular purposes for assessment. The first is using assessment in support of daily instruction and is named as ' assessment as integral to mathematics instruction', the second is large-scale assessment to influence what happens in the classroom and is named 'large-scale assessment'.

Rather than opposing different assessment approaches and different assessment formats we argue that different factors influence the choice for a particular assessment procedure. Some of these are of a paradigmatic character such as views about assessment; others are of a pragmatic character. These last factors refer to the purpose of assessment, the organizer or responsible, the learning objectives, the timing and the conditions in which the assessment take place.

2. Paradigmatic differences

These differences reflect: (1) different learning approach: behaviorism or constructivism (2) different methodological concerns about objectivity, reliability and validity.

In these days it is 'bon ton' to be a constructivist. Instruction methods such as anchored instruction and cognitive apprenticeship emphasize learning in meaningful contexts. Formative evaluation is essential. Subjective data are valuable. The typical formats used are constructed response assessment formats (f.i. performance assessment). Behaviorism at the other hand emphasizes the sequence of learning tasks, objective data and summative assessment. The purpose is to deliver a pre-specified body of knowledge. The typical formats used are selected response assessment formats (f.i. multiple-choice). Nevertheless also constructionism recognizes the importance of assessing easily retrievable fragments.

Methodological concerns accentuate these paradigmatic differences. Psychometric theory is well developed for selected response assessment formats. Objectivity, meaning clear-cut scoring without judgment of an examiner, is difficult to reach with constructed response assessment formats. This is also the case with reliability and validity. Face-validity, appearance of validity, is not sufficient. Hambleton (1996) discusses several interesting constructed response formats from a psychometric point of view. New IRT multidimensional models are in construction that can handle these forms of assessment.

Learning theory and methodological considerations influence the choice of a particular assessment format but in practice pragmatic considerations are decisive.

3. Pragmatic differences

Several questions help us in clarifying these pragmatic differences.

- Why are students assessed? What is the purpose of student assessment? Is it to check students ability in order to graduate students (to deliver diplomas) or to aid the students progress through the course? Is it to evaluate teachers, schools, school systems, and curricula or to enhance the quality of education by developing alternative learning routes? Is it used as an end or is it only a means to an end?
- Who organizes, who is responsible for the assessment? Is it the teacher, the researcher or the regional or national school system?
- Which learning objectives are assessed? Domain-specific or metacognitive skills? Basic skills or performance skills?
- When does the assessment take place? Before the course (or learning activity) or embedded in the daily learning activity and being integral to learning, or at specific moments during the course or after the course.
- What are the conditions in which the assessment takes place? These are the size of the group of students, amount of money and time, availability of resources etc.

Table 1. Summary

	Cicchitelli	Chance et al.	Starkings
WHY Purpose	Level of students ability	Quality of statistical education	Discussion of multichoice, group work and presentations. Emphasis on the diagnostic, formative and summative purpose of the assessment.
WHO Responsible	Team of item developers	Researchers	
WHAT Learning objectives	Knowledge and understanding of concepts and procedural skills of descriptive statistics	Visual understanding of the Central Limit Theorem	
WHEN Timing	After the course	Embedded in learning activity	
CONDITION	Large group Computers		

Starkings paper gives full account to pragmatic considerations and emphasizes methodological validity and reliability. Multichoice serves multiple purposes and is of good psychometric quality. Group work and presentations assess learning objectives that can't be assessed by multichoice, but suffers from poor psychometric quality.

Chances et al. paper emphasizes the use of assessment in research. They demonstrate how assessment can be used as a teaching tool. These authors rely on constructivist learning theory but nevertheless they use selected response format. The items are of the graphics-based type and are embedded in a contextual example.

Cicchitellis paper discusses selected response format and discusses validation referring to the traditional oral exams. The graphics-based item in his paper is difficult and does not discriminate between lower and higher achievers at the oral examination.

4. Conclusion

The three papers illustrate that assessment in statistics education reflects the different purposes and the different learning objectives of the statistics reform (Gal et al., 1997). Constructed response assessment is promising but a lot of research is still needed to cope with the methodological quality of this format.

REFERENCES

Gal, I. and J. Garfield (1997). Curricular goals and assessment challenges in statistics education. In I. Gal and J.B. Garfield (Eds.), The assessment challenge in statistics education (pp.1-13). Amsterdam: IOS Press.

Hambleton, R.K. (1996). Advances in assessment models, methods, and practices. In D.C. Berliner and R.C. Calfee (Eds.), Handbook of Educational Psychology (pp. 899-925). New York: Macmillan.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: steps toward a theory. In D.A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 661-683). New York: Macmillan.

RESUME

Les trois communications nous montrent que l'évaluation des acquis des étudiants en statistique doit être en accord avec le but et les objectifs de l'apprentissage en statistique. Tandis que l'utilisation des réponses construées ouvre des perspectives nouvelles dans l'évaluation de l'enseignement en statistique, les qualités méthodologiques de ce type ne sont pas encore mis à point. Il est argumenté qu'en faisant un choix de type de question avec réponse construée or réponse selectée, il y a deux genres de considérations qui interviennent. Elles sont de caractère paradigmatique comme théorie de l'apprentissage et qualités methodologiques et de caractère pragmatique comme but, responsable, objectif de l'apprentissage, le moment de l'évaluation et les conditions de l'évaluation des acquis.