Studies in Educational Evaluation 35 (2009) 193–199

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

Diagnosing skills of statistical hypothesis testing using the Rule Space Method

Seongah Im [a,*], Yue Yin [b]

[a] Department of Educational Psychology, University of Hawaii, 1776 University Avenue, Wist Hall 214, Honolulu, HI 96822, United States
[b] University of Illinois at Chicago, United States

ARTICLE INFO

Article history:
Received 18 August 2009
Received in revised form 15 December 2009
Accepted 17 December 2009

Keywords:
Cognitive diagnostic assessment
Rule Space Method
Statistical hypothesis testing
Educational evaluation

ABSTRACT

This study illustrated the use of the Rule Space Method to diagnose students' proficiencies in, skills and knowledge of statistical hypothesis testing. Participants included 96 undergraduate and, graduate students, of whom 94 were classified into one or more of the knowledge states identified by, the rule space analysis. Analysis at the level of proficiency groups showed that the critical difference, between low and medium proficiency groups was the understanding of statistical concepts and, knowledge while the critical skill discriminating the medium proficiency group from the high, proficiency group was to mange complex computational procedures. In addition, attribute profiles of, two students showed how students with the same total score can possess different strengths and, weaknesses.

Published by Elsevier Ltd.

Introduction

Researchers and educators have reached a consensus that a summative score does not improve students' learning (Bell & Cowie, 2001; Black & William, 1998; Nichols & Sugrue, 1999). Besides knowing a student's total score, it is important to know his or her diagnostic information with specific strengths and weaknesses so that teachers can provide appropriate feedback to students to help their learning. Unfortunately, it is often challenging for instructors to obtain diagnostic information about college students who are taking introductory statistics courses because these courses tend to have large class sizes and students from diverse educational backgrounds and preparation. This paper examines the use of the Rule Space Method in diagnosing students' knowledge and skills of statistical hypothesis testing using a statistics assessment at the collegiate level.

Statistics assessment

Statistics, especially at the introductory level of collegiate education, is an important subject and widely required for undergraduate and graduate students of many majors, such as education, sociology, psychology, and linguistics. Unfortunately, many students consider statistics one of the most difficult subjects (Schacht & Stewart, 1990), feel anxious about taking it, postpone enrolling in it until as late as possible, and regard undertaking statistics as negative and even identify it as a major threat to the attainment of their degrees (Onwuegbuzie, 2000).

Sometimes, statistics assessments may not greatly help students' statistics learning because they often focus on testing computational skills, knowledge isolated from reality, and the rote memory of formulas, vocabulary, and techniques (Garfield, 2003; Onwuegbuzie, 2000). These types of assessments tend to make students regard statistics as irrelevant and unimportant (Onwuegbuzie, 2000). Moreover, even if students calculate the correct answers, they are still unable to solve statistics problems in novel situations, adequately communicate statistical findings, effectively use statistical terminology in communication (Garfield, 1994), abandon misconceptions (Garfield, 1994; Garfied & Chance, 2000), apply obtained knowledge and skills in actual statistical analyses, and/or understand the structure and relationship among statistical concepts (Nikto & Lane, 1990).

To address these problems, many researchers have developed assessments to identify students' misconceptions (e.g., Capilla, 2007; Lipson, 2007), provide an instructive intervention (e.g., Hirsch & O'Donnell, 2001), and assess students' knowledge structures (e.g., Verkoeijen, Imbos, van de Wiel, Berger, & Schmidt, 2002). Researchers have also suggested that using computer software could enhance students' statistical understanding and problem solving (e.g., Konold, 1995) while identifying misconceptions and missing concepts and links in students' knowledge (e.g., Nikto & Lane, 1990).

In spite of improved assessment methods, it is often challenging for statistics instructors to address their students' individual needs and provide students with specific feedback for two reasons: (a) introductory statistics courses usually have large class sizes, and (b) the students have varying academic backgrounds and different

* Corresponding author. Tel.: +1 808 956 4299; fax: +1 808 956 6615.
E-mail address: seongahi@hawaii.edu (S. Im).

0191-491X/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.stueduc.2009.12.004

levels of preparation. The current study presents a possible solution to this problem using a psychometric model for cognitive diagnostic testing.

Rule Space Method

Researchers have developed multiple ways to enumerate strengths and weaknesses in individual performance. Such models include latent-class models (Maris, 1999), the Unified Model (DiBello, Stout, & Roussos, 1995), the Rule Space Model (Tatsuoka, 1985, 1991, 2009), the Attribute Hierarchy Model (Leighton, Gierl, & Hunka, 2004), the Partially Ordered Classification Model (Tatsuoka, 2002), and the General Diagnostic Model (von Davier, 2005, 2007).

Among these methods, the Rule Space Method (RSM; Tatsuoka, 1985, 1990, 2009) was designed to infer an individual student's mastery of latent skills and knowledge required to solve particular test items correctly on the basis of their observed item responses. The RSM uses students' item responses to infer their proficiency in certain knowledge, understanding, skills, and processes involved with solutions of the test items. Such knowledge, understanding, and skills are called *attributes* (Tatsuoka, 1990). The method characterizes those attributes by combining a cognitive analysis of a set of test items and psychometric procedures.

Specifically, the RSM consists of several steps. First, content experts identify specific attributes involved with a set of test items and incorporate these attributes in an incidence matrix, called a *Q-matrix*. The *Q*-matrix is a binary matrix composed of *k* columns and *i* rows where *i* rows represent item vectors and *k* columns represent attribute vectors. The entries in the *Q*-matrix indicate the incidence of item-attribute involvement (Tatsuoka, 1991). Fig. 1 illustrates a *Q*-matrix with five items and three attributes. Item 1 requires attributes A_1 and A_3 for a correct response, Item 2 requires attribute A_3, Item 3 requires attributes A_2 and A_3, and so on.

Second, the possible attribute patterns, called *knowledge states*, are generated from the *Q*-matrix. Given the example in Fig. 1, the total number of knowledge states generated is $2^3 = 8$. The knowledge states can range from the 'knowing-nothing-state' (0 0 0) to the 'knowing-all-state' (1 1 1). A knowledge state, for example (0 1 1) indicates knowing A2 and A3, but not A1. Then, the response patterns corresponding to the knowledge states are identified using a function of Boolean algebra (Tatsuoka, 1991). For example, the ideal response pattern corresponding to a knowledge state (0 1 1) is (0 1 1 0 1) because knowing A2 and A3 results in right answers to I_2, I_3, and I_5 according to the item-attribute involvement shown in the *Q*-matrix in Fig. 1.

Third, a two-dimensional Cartesian classification space, called *rule space*, is established with two axes of theta (θ) and zeta (ξ). Theta is the latent ability of an examinee defined in the two-parameter logistic item response theory (IRT) model. Zeta is an IRT-based measure of atypicality of the examinee's response patterns that indicates the tendency of a response pattern to answer difficult items correctly and easy items incorrectly (Tatsuoka, 1996). Next, the item responses corresponding to the knowledge states are projected into the rule space, and thus it determines the locations of the knowledge states in the rule space. Students' observed responses are also projected onto the locations of the knowledge states within the rule space.

Fourth, an examinee's knowledge states are determined by examining Mahalanobis generalized squared distances in the rule space between the examinee's coordinate and those of the knowledge states. The squared Mahalanobis distance follows a chi-square distribution with two degrees of freedom (Tatsuoka & Tatsuoka, 1987). Knowledge states within a cutoff of the distance given the examinee's location at the rule space are accepted as the examinee's plausible knowledge states. In addition, the Bayes' decision rule for minimum error is also applied because the distance measure itself does not provide misclassification probabilities (Tatsuoka & Tatsuoka, 1987). Mastery of an attribute for an individual examinee is expressed in probabilistic form called *attribute mastery probability* (AMP). It is a combination of the knowledge states weighted by their posterior probabilities calculated from the Bayes' decision procedures. When the examinee's mastery probabilities of the specified attributes are calculated, the cognitive diagnosis for the examinee is complete.

The RSM has been proven to be especially effective when applied to a well-defined domain, in which a certain concept or theory is defined in a logical way using a set of rules or procedures in an unambiguous way. When a domain is well defined, it is straightforward to specify a set of knowledge and skills involved with the test items in the domain (Nikto, 1984).

The applicability of the RSM in such domains of secondary education has been examined using mixed number subtraction (Birenbaum & Shaw, 1985), linear algebraic equation (Birenbaum, Kelly, & Tatsuoka, 1993), and multiplication and division with exponents (Birenbaum & Tatsuoka, 1993). Studies on university students demonstrated as well that the RSM was useful for characterizing students' proficiencies in more complex domains, such as an architecture assessment (Katz, Martinez, Sheehan, & Tatsuoka, 1998) and a reading comprehension assessment (Buck, Tatsuoka, & Kostin, 1997). Although the RSM has been used in the assessment of many content subjects, it has not been applied in statistics assessment.

Using the Rule Space Method in statistics assessment

We propose to apply the RSM in a classroom assessment of statistics for the following reasons. First, statistics is a well-defined domain in which explicit rules in problem solving can be clearly identified and formulated. This feature makes the application of the RSM in statistics assessment appropriate and possible.

Second, the RSM can be built on existing statistics assessment. There were many studies in which the RSM was productively applied to large-scale assessments that had not designed with a diagnostic purpose in mind. Such studies used the mathematics tests of Trends in International Mathematics and Science Study (TIMSS) (e.g., Birenbaum, Tatsuoka, & Yamada, 2004; Tatsuoka, Corter, & Tatsuoka, 2004) and the quantitative test of Graduate Record Examinations (GRE) (e.g., Tatsuoka & Boodoo, 2000). Because the RSM does not necessarily require designing special assessments to match the method, this makes the application of the RSM in statistics assessment more feasible and flexible.

Third, the RSM can provide timely and efficient diagnostic information, which makes the application of the RSM particularly beneficial in introductory statistics classes. As discussed earlier, introductory statistics classes tend to have large class sizes. The RSM can help instructors obtain diagnostic information efficiently from many students in a large class because the RSM simply

$$\begin{array}{c} \\ I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

Fig. 1. An example *Q*-matrix composed of five items (I_1–I_5) and three attributes (A_1–A_3).