Learning and Individual Differences 21 (2011) 392–402

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Measuring mathematical ability needed for "non-mathematical" majors: The construction of a scale applying IRT and differential item functioning across educational contexts

Silvia Galli *, Francesca Chiesi, Caterina Primi

Department of Psychology, University of Florence, Italy

ARTICLE INFO

Article history:
Received 8 November 2010
Received in revised form 11 March 2011
Accepted 14 April 2011

Keywords:
Mathematical ability
Educational assessment
Item response theory
Differential item functioning
Cross-country invariance

ABSTRACT

Given that basic mathematical ability is a requirement to succeed in "non-mathematical" majors, e.g. degrees for Psychology, Education, and Health Sciences with compulsory introductory stats courses, assessing this ability can be useful to promote achievement. The aim of the present study was to develop a scale to measure the mathematical ability deemed necessary to successfully complete introductory statistics courses. The Item Response Theory was applied to construct a scale for accurately measuring low levels of ability, i.e. levels under which examination attainment should be prevented. Mathematical basics were identified and a pool of items was created. Items were calibrated under the 2PL model and the test information function was investigated. The equivalence of the developed scale across national (Italy) and international (Spain) contexts was assessed analyzing Differential Item Functioning (DIF). Finally, the validity was studied for both the Italian and the Spanish versions of the scale. Advantages of the scale for international research and teaching are discussed.

1. Introduction

Students attending undergraduate programs as degrees for Psychology, Education, and Health Sciences encounter difficulties in courses with mathematical contents, i.e. introductory stats courses that are compulsory in their programs. As a result, these courses might delay the completion of the degree or have as a consequence the non-completion the degree.

Even if these courses do not require sophisticated mathematical abilities, basic mathematical ability is a necessary prerequisite to successfully complete them. In detail, operations, i.e., addition, subtraction, multiplication, division with fractions, and exponentiation, are required in descriptive and inferential procedures (e.g., to compute the standard deviation, or the *t* o *z* values); first order equations are deemed necessary in the standardization procedure and in the regression analysis; the set-theory principles, fractions and decimal numbers from 0 to 1 were considered necessary to deal with probability (e.g., to compute the occurrence of an event, or to interpret probability values); establishing relations between numbers from −1 to 1, and the concept of absolute value were considered necessary for drawing conclusions in hypothesis testing (e.g., to compare the computed and the critical values in order to decide if the null hypothesis has to be accepted or rejected).

Several studies (e.g. Chiesi & Primi, 2010; Harlow, Burkholder, & Morrow, 2002; Nasser, 2004; Schutz, Drogosz, White, & Distefano, 1998; Tremblay, Gardner, & Heipel, 2000) found a positive effect of previous acquired mathematical competence on achievement in introductory statistics course and, moreover, basic mathematical competence was found to be related to attitudes and anxiety toward statistics, which are deemed important factors influencing college students' learning of statistics (for a review, Zieffler et al., 2008). Indeed, previous experiences in mathematics might affect attitude toward statistics (Carmona Marquez, Martìnez, & Sànchez, 2005; Dempster & McCorry, 2009; Gal, Ginsburg, & Shau, 1997), which in turn affects achievement. Concerning anxiety, students with low mathematical ability frequently experienced high levels of anxiety towards statistics that may impair learning and performance by interfering with students' ability to receive, concentrate on, and encode statistical issues (Baloglu, 2003; Hong & Karstensson, 2002; Onwuengbuzie, 2003; Zeidner, 1991).

Given that basic mathematical ability is a fundamental requirement to succeed in "non-mathematical" majors, assessing this ability can be useful to promote performance. Specifically, assessing mathematical competence can be useful to identify students with a considerable impairment in this domain that might be a severe obstacle to the attainment of the course. Once identified, these students could be aided in improving their competences, and consequently helped to pass the exams. Till now, for assessing mathematical ability math grades at high school levels were used (Onwuegbuzie, 2003; Nasser, 2004; Sorge & Schau, 2002), or tests

* Corresponding author at: Department of Psychology, via di San Salvi 12-Padiglione 26, 50135 Firenze, Italy. Tel.: +39 0556237846; fax: +39 0556237867.
E-mail address: silvia_galli@hotmail.it (S. Galli).

1041-6080/$ – see front matter
doi:10.1016/j.lindif.2011.04.005

developed for instructional use or for the purpose of research were employed. For instance, Tremblay et al. (2000) presented a 10-item test including rounding, computation, and transformation of percentage to decimal, basic graph interpretation and solving worded problems, Schutz et al. (1998) administered a 30-item math test including the use of exponents, multiplication, division of fractions, and other basic algebra applications, Harlow et al. (2002) used a 30-problem math skills test, adapted from Gravetter and Wallnau (1996), to measure the initial levels of students' quantitative understanding. However, these measures have some limitations: on one hand, grades are not comparable because they depend on different scholastics background with different level math courses; on the other hand, psychometric properties of tests developed for the purpose of specific researches have not been investigated or reported. In the same way, math subscales of standardized tests (e.g., *Scholastic Aptitude Test* and *American College Testing*) might lack the specificity to assess mathematical ability deemed necessary for stats course included in "non-mathematical" majors.

Starting from these premises, the aim of the present study was to develop a scale to measure the mathematical ability deemed necessary to successfully complete introductory statistics courses. Indeed, answering current questions about how statistical education and teaching should develop to ensure sound statistical understanding and usage in a wide range of disciplines is a very current topic around the world (e.g., Garfield & Ben-Zvi, 2009). This implies there is a need for research that informs about what works and what is relevant in different communities and, even if different communities might differ widely, an accessible research literature is clearly essential for promoting statistics education from an international perspective (see 2010 International Conference on Teaching Statistics, ICOTS8). In particular, an international perspective requires instruments employable across different countries, that is, tools that are able to obtain comparable results for international research collaboration.

In particular, the purpose of this research was to develop a scale to accurately measure low levels of mathematical ability. The information obtained from the scale, that is the identification of students with relevant difficulties in this domain, should be useful to improve achievement and prevent failures since those students could be supported from the first day of the course with specific mathematics training courses.

The scale was developed using item response theory (IRT) because it offers a different value of test precision for each specific level of underlying latent variable being measured and it does not assume that a single estimate of reliability, and corresponding standard error of measurement, is sufficient to describe precision of measurement over all levels of ability (Lord, 1980). In other words, IRT is appropriate for developing instruments aimed at accurately measuring a specific level of the assessed ability (Embreston & Reise, 2000).

In order to develop the scale, as a first step, mathematical basics to successfully complete introductory statistics courses were identified, and a pool of items was created (*Construct Definition and Item Construction*). Then items were calibrated and the test information function, i.e., the reliability of the scale for different levels of ability, was investigated (Study 1). Afterward, it was investigated the measurement equivalence of the developed scale across national and international contexts (Study 2). In order to check the invariance of the scale across national contexts, it was administered to samples of Italian students enrolled in different psychology faculties. To test the invariance across international contexts the scale was administered to Spanish psychology students which, as previously documented (Carmona Marquez, 2004; Carmona Marquez et al., 2005), encounter difficulties in dealing with statistics. These students attended an introductory statistics course that covers the same topics of the Italian ones. For both national and international contexts the measurement equivalence of the scale was assessed analyzing the *Differential Item Functioning* (DIF) for each item across contexts using IRT. Within the IRT framework, an item is said to exhibit DIF if respondents of two different groups who have equal level of ability being measured do not have the same probability of answering correctly (Lord, 1977). More precisely, when an item parameter (difficulty, discrimination or guessing) differs across groups, an item displays DIF.

Finally, the validity of the developed scale was studied for both the Italian and Spanish versions (Study 3). Relying on some researches on statistics achievement (for a review, Zieffler et al., 2008), we expected to find that mathematical ability would be positively related to attitudes toward statistics, and inversely related to anxiety toward statistics. Moreover, we investigated the relationship between mathematical ability and achievement in statistics expecting to find some predictive utility of the mathematical ability measure.

2. Construct definition and item construction

In order to define the construct of mathematical prerequisites necessary to successfully complete introductory statistics courses the contents of these courses were examined. Specifically, we examined introductory statistics course curricula and contents of textbooks. Introductory statistics courses are designed to prepare students to handle, use and interpret research or statistical data in their field of study. Students are first introduced to descriptive statistics: frequency distribution, several measures of central tendency and variability, and relationship measures. This is followed by presentation of some theoretical and procedural aspects of inferential statistics such as sampling distribution, hypothesis testing, and confidence interval estimation (e.g. Agresti & Finlay, 1999; Ercolani & Perugini, 1997; Primi & Chiesi, 2005; Welkowitz, Cohen, & Ewen, 2006).

Two teachers of statistics identified the necessary mathematical basics to successfully complete the introductory statistics courses. In detail, addition, subtraction, multiplication, division with fractions, and exponentiation, required in descriptive and inferential procedures; the set-theory principles, fractions and decimal numbers from 0 to 1 necessary to deal with probability; first order equations necessary in the standardization procedure and in the regression analysis; relations between numbers from −1 to 1, and the concept of absolute value for drawing conclusions in hypothesis testing; finally, some basics of probability were also included.

In order to operationalize these prerequisites an initial pool of 48 items was developed. Each item presented a multiple choice question (one correct among four alternatives). Items were based on existing materials such as mathematics textbook tasks. After the formulation, items were evaluated at the qualitative level. Two teachers of mathematics analyzed their contents and checked the wording to find out if they were suitable and intelligible. As a result of this analysis, 6 items were modified and 18 items were removed. More precisely, items with one or more ambiguous response alternatives were adapted, whereas those with both ambiguous stems and response alternatives were removed. A 30 item scale was finally obtained in which items covered a broad range from low to medium difficulty levels[1] according to the purpose of measuring basic mathematical ability. For this reason, the scale was named *Prerequisti di Matematica per la Psicometria* (PMP) [Math Prerequisites for Introductory Statistics] (Appendix A).

3. Study 1

This study aimed to calibrate the scale, i.e., to estimate the item parameters, and to examine its reliability for different levels of ability.

[1] In a preliminary study (Galli, Chiesi & Primi, 2008) the difficulty of the 30 items was checked applying the Rasch Simple Logistic (RSL) model (Rasch, 1960).