Item Response Theory for Optimal Questionnaire Design

Giuseppina Lotito[1], Giuseppe Pirlo[2]

[1] IISS "Gorjux-Tridente", Baru, Italy
giuseppina.lotito@istruzione.it
[2] University of Bari "Aldo Moro", Italy
giuseppe.pirlo@uniba.it

Student assessment is one of the most critical aspects related to web-based learning systems. In this field, the use of on-line questionnaires - based on multiple-choice items - is one of the most widespread approaches.

This paper presents a new technique for automatic design of optimal questionnaires that uses a Genetic Algorithm for multiple-choice item selection, according to the Item Response Theory.

The experimental results, carried out on both simulated and genuine data, confirm the effectiveness of the new approach, that is able to adapt questionnaire design to the abilities of a given set of students.

1 Introduction

Accurate assessment of learning processes is one of the key aspects in the knowledge society. Actually, learning process assessment concerns not only the effectiveness of a learning activity on a student's skill, but it also provides feedback to course designers and instructors that can then improve learning products and services, as well as determine the most effective organization strategies for learning processes (Lan *et al.*, 2011; Romero *et al.*, 2010).

Recently, along with the spread of learning systems based on Information and Communication Technologies (ICT), computer-based student assessment is gaining attention since it is considered a fundamental service of Learning Management Systems (Alemung *et al.*, 2011; Dimauro *et al.*, 2003; Romero *et al.*, 2008; Greco *et al.*, 2006b). Student assessment has both formative and summative purposes. formative assessment takes place several times during a course and aims to focus on cognitive, social, and motivational aspects of learning. Summative assessment occurs at the end of a course and aims to evaluate the cognitive aspects of learning. It is designed and conducted by the teacher and does not take into consideration the learning process (Strijbos *et al.*, 2011; Dimauro *et al.*, 2006; Pirlo *et al.*, 2008; Impedovo *et al.*, 2011; Greco *et al.*, 2006a).

Whatever purpose is considered, student assessment is rightly considered a fundamental part of the learning process and several types of computer-based systems for a student's assessment have been proposed (Lan *et al.*, 2011; Romero *et al.*, 2010). The most popular systems use questionnaires based on multiple-choice items, in which students are asked to select the best possible answer from the choices provided on a list (Kuechler *et al.*, 2003). Multiple-choice items offer the possibility of generating useful data that can provide a better understanding of the learning process (Romero *et al.*, 2009; Impedovo *et al.*, 2006). For example, students' questionnaire data have been successfully used for individual target analysis (Yamanishi *et al.*, 2001), for discovering the individual needs of the students (Pechenizkiy *et al.*, 2008), for providing personalized learning suggestions (Chu *et al.*, 2006), as well as for discovering rule patterns (Chen *et al.*, 2009). In addition, multiple-choice items can be easily integrated into computer-based assessment systems since they support fast automatic evaluation and reuse (Kuechler *et al.*, 2003; Romero *et al.*, 2009). Unfortunately, little attention has been devoted so far to the questionnaire design process. In fact, the design of a questionnaire is a complex task since it requires the selection of the set of items most advantageous for assessing the skill level of a student (Lan *et al.*, 2011).

This paper presents a new approach for optimal questionnaire design. The approach first uses the Item Response Theory to estimate the item difficulty for

a given student class with specific abilities. Successively the approach considers a genetic algorithm to determine the best set of items to be included in the questionnaire for that specific class of students. The basic idea of this paper is that the effectiveness of a questionnaire is strictly connected to the students for which the questionnaire is designed. In other words, a questionnaire is an entity that must be tailored according to the specific characteristics of the class of students that must be assessed.

The organization of the paper is the following. Section 2 presents the problem of questionnaires design for student assessment, based on Item Response Theory. Section 3 presents the genetic algorithm used for automatic questionnaire design. Section 4 presents the experimental results. Section 5 reports the conclusion.

2 Item Response Theory for Automatic Item Selection

Item Response Theory is a well-known paradigm of psychometrics that is based on the consideration that responses to a set of items can be explained by the existence of one or more latent traits, named abilities (van der Linen *et al.*, 1997). A latent trait, generally represented by the θ symbol, is conceptualized as a quantitative trait and is generally scaled to have a mean of zero and a standard deviation of one. A main objective in item response modelling is to characterize the relation between θ and the probability of item endorsement. This relation is typically referred to as the Item Characteristic Curve (ICC) and can be defined as the (nonlinear) regression line that represents the probability of endorsing an item (or an item response category) as a function of the underlying trait (Fraley *et al.*, 2000). Though a complete description of the Item Response Theory is beyond the scope of this paper, the interested reader can find a comprehensive analysis in the literature (van der Linen *et al.*, 1997; Fraley *et al.*, 2000). For the purpose of this work, we have taken into consideration the Two-Parameter Logistic Model (2PLM) (Birnbaum *et al.*, 1968). In this case, letting $T=\{t_1, t_2,\dots, t_j\dots, t_M\}$ be a set of items, the probability that an individual with trait level θ_i will endorse item t_j is defined as a function of two item parameters: the item difficulty parameter β_j and the item discrimination parameter α_j (Birnbaum *et al.*, 1968):

$$P_j(\theta_i) = \frac{1}{1+e^{-\alpha_j(\theta_i-\beta_j)}}$$

(1)

where: the difficulty parameter β_j represents the level of the latent trait ne-

cessary for an individual to have a 50% probability of endorsing the item; the item discrimination parameter α_j represents an item's ability to differentiate between people with contiguous trait levels.
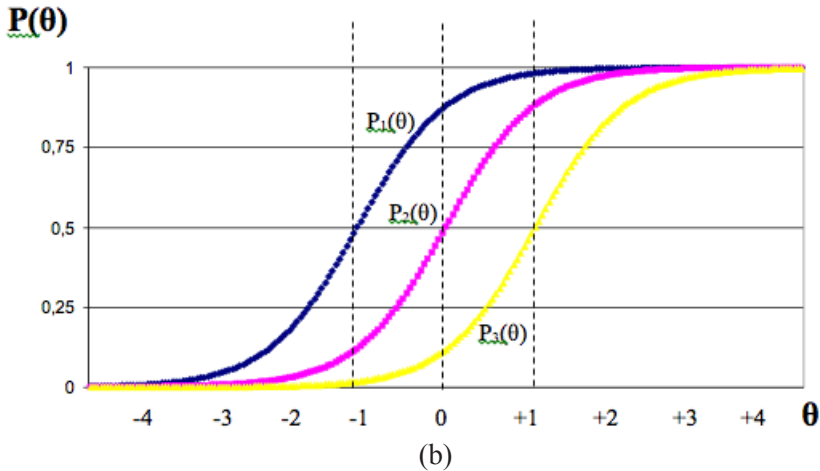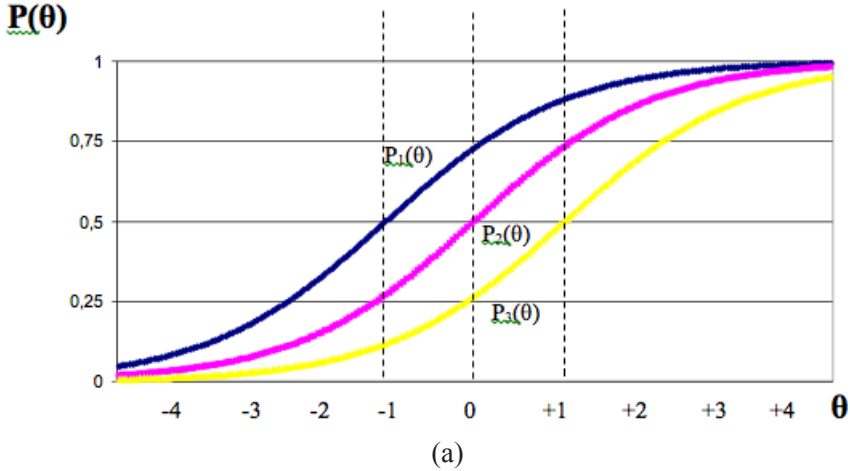

(a)

(b)

Fig. 1. Item Characteristic Curves (ICCs)

Figure 1a shows the ICCs for three items with $\alpha_1=\alpha_2=\alpha_3=1$ and $\beta_1=-1$; $\beta_2=0$; $\beta_3=+1$. Figure 1b shows the ICCs for three items with $\alpha_1=\alpha_2=\alpha_3=2$ and $\beta_1=-1$; $\beta_2=0$; $\beta_3=+1$. It is worth noting that items are not equally informative across the entire range of the trait θ. In fact, an item yields the most information when θ^i equals β_j. In other words, items are most informative when the item difficulty

parameter is perfectly matched to the person's trait level (van der Linen *et al.*, 1997; Birnbaum, 1968). Furthermore, in the context of Item Response Theory, an item is considered difficult if a high level of ability or knowledge is required to answer it correctly. Only individuals with a high degree of knowledge will be able to answer the difficult items, and almost everyone will be able to answer the easy items. Therefore, the difference $P_j(\theta_{max})-P_j(\theta_{min})$ can be used to estimate the extent to which item tj is valuable to assess students in the range $[\theta_{min}, \theta_{max}]$: the greater the difference $P_j(\theta_{max})-P_j(\theta_{min})$ the better the item t_j. Figure 2 shows the ICCs of two items t_1 and t_2. In this case, the results indicated that t_1 is better than t_2 for assessing the students in the range $[\theta_{min}, \theta_{max}]$, since $P_1(\theta_{max})-P_1(\theta_{min})$ > $P_2(\theta_{max})-P2(\theta_{min})$.
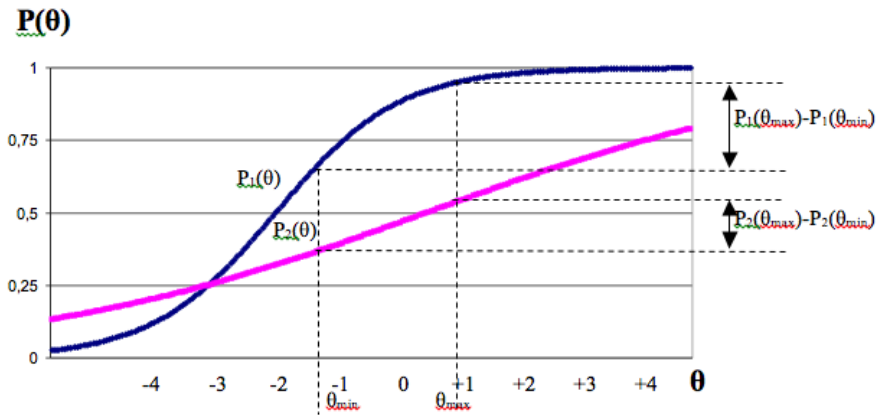

Fig. 2 - Student Estimation by ICCs

Of course, when a set of independent items $T=\{t_1, t_2,\ldots, t_j\ldots, t_M\}$ is considered, the probability that an individual with trait level θi will be able to endorse all items is defined as (Fraley *et al.*, 2000):

$$P^T(\theta_i) = \prod_{j=1}^{M} P_j(\theta_i)$$

(2)

3 Questionnaire Design by Genetic Algorithms

From a broad set of M items, this new technique for automatic questionnaire design selects the most profitable subset of N items (N<M) to be included in the questionnaire for a certain category of skills. More precisely, let $T=\{t_1, t_2,\ldots,$

$t_j\ldots, t_M$} be the set of M items available, and $S=\{s_1, s_2,\ldots, s_i\ldots, s_N\}$ the set of N students under consideration. Furthermore, let θi be the i-th student trait ability level, for $i=1,2,\ldots,N$. Following the considerations in the previous section, the questionnaire design is here expressed as an optimization problem in which the subset of items $Q=\{t_{ip} \mid p=1,2,\ldots,P$ with $(1\leq_{ip}\leq M$ and $_{ip}\neq_{iq}$ for $p\neq q)\}$, must be selected to maximize the fitness function:

$$F(Q) = P^Q(\theta_{max}) - P^Q(\theta_{min})$$

(3)

with:
- θ_{max}=maximum skill value for the set of students
  (i.e. $\theta_{max}=\max\{\theta_i \mid i=1,2,\ldots,N\}$) (4a)
- θ_{min}=minimal skill value for the set of students
  (i.e. $\theta_{min}=\min\{\theta_i \mid i=1,2,\ldots,N\}$) (4b)

Thus the optimization process has to select - from the set of items T - the subset more suitable for investigating the latent abilities of the set of students belonging to the skill range $[\theta_{min}, \theta_{max}]$.

In this paper, a binary-coded genetic algorithm is used to solve the optimization problem in eq. (3), since genetic algorithms – as widely discussed in the literature - have potential for solving non-linear optimization problems, in which the analytical expression of the object function is not known. Moreover, genetic algorithms are able to depart from local optima, unlike deterministic, gradient–based optimization methods, which tend to converge towards local extrema of the object function (Michalewicz, 1996). A complete description of genetic algorithms is not provided here, but any reader who is interested can find excellent survey papers in the literature (Goldberg, 1989). The following describes the genetic algorithm used in our approach (Baeck, 1996).

I) The initial – population $P_{op}=\{\Phi_1, \Phi_2,\ldots,\Phi_к,\ldots,\Phi_{Npop}\}$ of random individuals was created. In our tests Npop has been set to 20 since some preliminary experiments have shown Npop = 20 is a good trade-off between convergence speed of the genetic algorithm and its capability to escape from local extrema. In our approach, each individual was a questionnaire Q and it was represented by a vector $\Phi_k=<h_1,h_2,\ldots,h_j,\ldots,h_M>$, where each gene hj was a Boolean value:
- $h_j=0$ means that j-th item of T (i.e. the item t_j) was not included in Q;
  (5a)
- $h_j=1$ means that j-th item of T (i.e. the item t_j) was included in Q.
  (5b)

For example, let T={t_1, t_2, t_3, t_4, t_5, t_6, t_7, $t8$} be the set of items. The individual Φ_k=<h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8>= <1,0,0,1,1,1,0,1> represented the questionnaire Q={t_1,t_4,t_5,t_6,t_8}.

Since P items must be included into the questionnaire Q, the following normalization procedure was performed for each individual Φk:

a) Compute $$P' = \sum_{j=1}^{M} h_j \qquad (6a)$$

b) If P'>P then select randomly (P'-P) genes equal to 1 and set them to 0
 if P' < P then select randomly (P-P') genes equal to 0 and set them to 1. (6b)

After normalization, the fitness function was computed for each individual Φ_k of the population, according to eq. (3).

II) From the initial - population, the following four genetic operations were used to generate the new populations of individuals:

i) Individual Selection. In the selection procedure N_{pop}/2 random pairs of individuals were selected for crossover, according to a *roulette-wheel* strategy. This associates a selection probability to each individual. The higher the fitness function of the individual, the higher the selection probability (Baeck, 1996).

ii) Crossover. Crossover is a probabilistic process that exchanges information between two parent individuals selected for crossover:

<$h^a_1,h^a_2,.,h^a_{v-1}, h^a_v,..,h^a_M$> and <$h^b_1,h^b_2,...,h^b_{v1}, h^b_v,...,h^b_M$>, $\qquad (7a)$

for generating two offspring individuals of the next generation:

<$h^a_1,h^a_2,.,h^a_{v-1}, h^a_v,..,h^a_M$> and <$h^b_1,h^b_2,...,h^b_{v1}, h^b_v,...,h^b_M$>. $\qquad (7b)$

In our approach, a one-point crossover was used (Baeck, 1996). In this case, for each pair of individuals selected for crossover, a random integer v ($1< v \leq$ M) was chosen and the child individuals in (7b) were generated as follows:

- $h^a_s = h^a_s$ and $h^b_s = h^b_s$, if $v < v$; $\qquad (8a)$
- $h^a_s = h^b_s$ and $h^b_s = h^a_s$, if $v \geq v$. $\qquad (8b)$

For instance, if the individuals
$$<h^a_1, h^a_2, h^a_3, h^a_4, h^a_5, h^a_6, h^a_7, h^a_8>=<1,0,1,1,0,0,1,1>$$
and
$$<h^b_1, h^b_2, h^b_3, h^b_4, h^b_5, h^b_6, h^b_7, h^b_8>=<1,1,0,1,0,1,1,0>$$

are considered, for s=5 the offspring are

$$<h^a_1, h^a_2, h^a_3, h^a_4, h^a_5, h^a_6, h^a_7, h^a_8>=<1,0,1,1,0,1,1,0>$$
and
$$<h^b_1, h^b_2, h^b_3, h^b_4, h^b_5, h^b_6, h^b_7, h^b_8>=<1,1,0,1,0,0,1,1>$$

iii) Mutation. A mutation operator can be applied to change some of an individual's genes. Mutation is used to prevent falling genetic algorithm into local extreme. A uniform mutation operator was applied in this study. Let $\Phi_k=<h_1,h_2,...,h_M>$ be an individual, the uniform mutation operator changed (inverted) each gene of the individual according to a mutation probability, *Mut_prob* (*Mut_prob*=0.02 in our tests). After mutation, in order to ensure that each questionnaire contained a number of items equal to P, the normalization procedure performed by eqs 6a,b was then applied to all individuals Φ_k, k=1,2,…,Pop.

iv) Elitist Strategy. In our approach, an elitist strategy was adopted. From the Npop individuals generated by the above operations, one individual was randomly removed and the individual with the maximum fitness in the previous population was added to the current population (Baeck, 1996).

Operations (i),(ii),(iii),(iv) were then repeated until N^{iter} successive populations of individuals were generated (N^{iter}=50 in our tests). When the process stopped, the optimal questionnaire was obtained by the best individual of the last-generated population.

4 Experimental Results

In order to evaluate the new technique for optimal questionnaire design, the experimental tests were carried out on both simulated and real data. Both experiments included two steps. In the preliminary step student models (i.e. the trait ability level of each student) were estimated. In the test step the optimal questionnaire was designed for the specific set of students under consideration. For the experiments on simulated data, a suitable software was developed to generate the set of MT*N random responses simulating the answers of N of students to a set of MT items. Figure 3 shows a screenshot of the software interface. In the preparation step, after data simulation, the ICC of each item

was evaluated using the 2PLM model and the trait ability level of each student was computed. For the purpose, the Marginal maximum likelihood estimation was considered, where the hidden student variables are chosen to maximize the likelihood of the data, according to the approach proposed in the literature (Bock & Aitkin, 1981)

Finally, the skill range of the set of students $[\theta_{min}, \theta_{max}]$ was determined. In the test step, a new set of M items named Full Set (FSM) was generated and the optimal questionnaire T*P could be defined by automatically picking out the optimal subset of P items from FSM, for the given set of simulated students with a range equal to $[\theta_{min}, \theta_{max}]$.
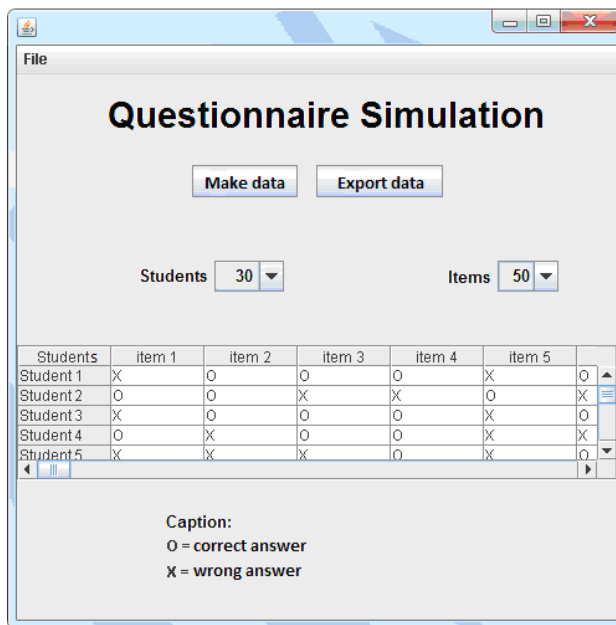

Fig. 3 -The questionnaire simulation procedure (screenshot)

Figure 4 shows the experimental results obtained with the simulation procedure. In this case, we considered N=30 students and MT=100 items. Successively, the ability of each student was estimated according to the approach of Bock and Aitkin (Bock & Aitkin, 1981) and the skill range $[\theta_{min}, \theta_{max}]$=[2.22, 3.27] of the student set was determined. The test step was carried out using the questionnaire FSM of M items (M=50 in our test) and other questionnaires obtained by selecting the optimal subset $T*_p$ of P items out of M (P=5,10,15 in our test). In particular, Figure 4a shows the result from the questionnaire of M=50 items. Figure 4b,c,d show the results from the questionnaires of M=5,

M=10 and M=15 items, respectively. Of course, in order to evaluate the effectiveness of the proposed approach, the ability estimated when using the optimal questionnaire T^*_P was compared with the average ability determined when using the random-generated questionnaires of P items, where item selection was performed randomly. In particular, each value T^{rnd}_P reported in Figure 4 is the average ability calculated when taking into account 10 questionnaires, each one realized by selecting P random items from FSM.

Similarly, Figure 5 presents the result on real data, for N=34 students. In this case, the skill range of the set of students [θ_{min}, θ_{max}] was equal to [1.67, 3.05]. A Full Set of M=60 items was provided by the teacher and the optimal questionnaire T^*_P was defined by automatically picking out from FSM the optimal subset of P items, P=5, 10, 15. In particular, Figure 5a shows the result from the questionnaire of M=50 items. Figure 5b,c,d show the results from the questionnaires of M=5, M=10 and M=15 items, respectively. Therefore, figure 5 shows the ability levels estimated through the FSM, T^*_P and T^{rnd}_P.
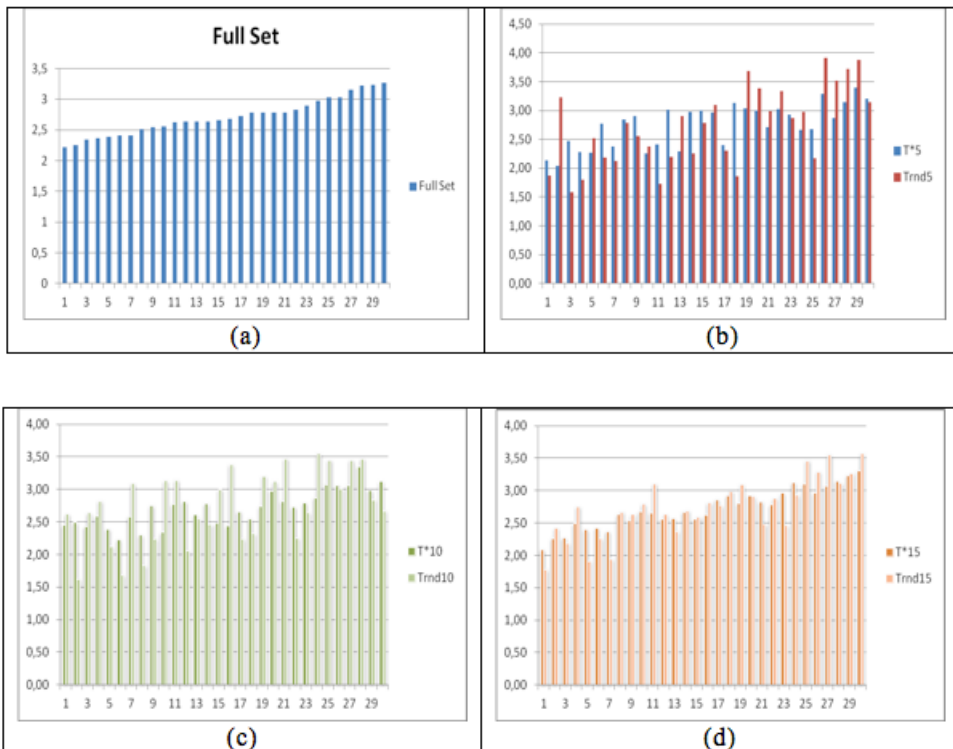

Fig. 4 - Experimental Results (Simulated Data)

Figure 5. Experimental Results (Real Data)

Now, in order to estimate the effectiveness of the questionnaires for student assessment we considered the following measures:

- A_FSQ(i) the ability of the i-th student estimated through the Full Set questionnaire FSM of M items;
- A_T*$_p$(i) the ability of the i-th student estimated through the optimal questionnaire T*$_p$ of P items;
- A_T$^{rnd}_p$(i) the ability of the i-th student estimated by averaging the abilities determined through 10 random-generated P items questionnaires.

Hence the accuracy of T*$_p$(i) and T$^{rnd}_p$(i) to assess student ability was estimated, respectively, by the standard deviations:

$$SD(FSQ\_T^*_p) = \sqrt{\sum_{i=1}^{N}\left[A\_FSQ(i) - A\_T^*_p(i)\right]^2}$$

(9a)

$$SD(FSQ\_T^{rnd}{}_p) = \sqrt{\sum_{i=1}^{N}\left[A\_FSQ(i) - A\_T^{rnd}{}_p(i)\right]^2}$$

(9b)

It is worth noting that the standard Eq. (9a) shows the extent to which the assessments carried out by the optimal subsets of P items (for different values of P) are different from the assessment obtained when all items are considered. Therefore, the lower SD(FSQ_T*$_p$) the most effective is assessment performed though the optimal subset of items. Similarly, in Eq. (9b), the standard deviation shows to what extent the assessments carried out by using random subsets of P items (for different values of P) differs from the assessment obtained when all items are considered. Of course, the comparison between SD(FSQ_T*$_p$) and SD(FSQ_T$^{rnd}{}_p$) reported in Table I provides a useful information about the capability of the proposed approach in selecting optimal subsets of items for questionnaire design, able to assess students more precisely than using randomly selected items.

	P	SD(FSQ_T*$_p$)	SD(FSQ_T$^{rnd}{}_p$)
	5	0.25	0.52
Simulated Data	10	0.17	0.40
	15	0.10	0.29
	5	0.20	0.51
Real Data	10	0.14	0.35
	15	0.06	0.24

More precisely, applying eqs. (9) to the results on simulated data in Figure 4 it follows that, for P=5, SD(FSQ_T*$_5$)=0.25 and SD(FSQ_T$^{rnd}{}_5$)=0.52. Therefore, the assessment accuracy of the optimized questionnaire T*5 was found to be superior to the accuracy of Trnd5 by 51.9%, on average. Similar results were also found for P=10, SD(FSQ_T*$_{10}$)=0.17 and SD(FSQ_T$^{rnd}{}_{10}$)=0.40. In this case the accuracy of the optimized questionnaire outperformed the accuracy of the random questionnaires by 57.5%, on average. In addition, when the optimized questionnaire was used for P=15, the standard deviation reduced by 65.5% in comparison with the standard deviation of random questionnaires, on average. In fact, in this case, SD(FSQ_T*$_{15}$)=0.10 and SD(FSQ_T$^{rnd}{}_{15}$)=0.29.

Also concerning the tests on real data (reported in Figure 5), the experimental results confirmed the evidence obtained using the simulated data. In particular, our results found that P=5, SD(FSQ_T*$_5$)=0.20 and SD(FSQ_T$^{rnd}{}_5$)=0.51. This means that, when the optimized questionnaire was used, the standard devia-

tion in the assessment of student ability reduced by 60.3% with respect to the standard deviation obtained by random questionnaires, on average. Similarly, for P=10 we found: $SD(FSQ\_T^*_{10})=0.14$ and $SD(FSQ\_T^{rnd}_{10})=0.35$. Therefore, the optimized questionnaire could be seen to reduce the standard deviation by 57.7% when compared to the random questionnaires, on average. Also for P=15, when the optimized questionnaire was used, the standard deviation reduced by 73.6% with respect to the average standard deviation of the assessment obtained by the random questionnaires. In fact, in this case, $SD(FSQ\_T^*_{15})=0.06$ and $SD(FSQ\_T^{rnd}_{15})=0.24$.

Conclusion

Multiple-choice item questionnaires are a widespread approach used for student assessment in web-based learning systems. In this domain, the problem of optimal questionnaire design is still open.

This paper addresses the problem of questionnaire design for student assessment and presents a new technique for adaptive questionnaire design based on Item Response Theory. Therefore, the aim of this work is twofold. First, the problem of optimal questionnaire design is considered as an optimization problem. Second, a genetic algorithm is proposed for optimal questionnaire design and its effectiveness is demonstrated. The algorithm automatically selected the best set of items for the specific range of ability of the students under consideration.

The experimental results, carried out on both simulated and genuine data, confirm the effectiveness of the new approach, that is able to adapt questionnaire design to the abilities of a given set of students.

REFERENCES

Amelung M., Krieger K., Rosner D. (2011), *E-Assessment as a Service*, IEEE Trans. on Learning Technologies, Vol. 4, No. 2, pp. 35-46.

Baeck T. (1996), *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolution Programming, Genetic Algorithms*. New York: Oxford Univ. Press.

Birnbaum A. (1968), *Some latent trait models and their use in inferring an examinee's ability*, Statistical theories of mental test scores, F. M. Lord and M. R. Novick (eds.), Addison-Wesley, pp. 397-472.

Bock, R. and Aitkin, M. (1981), *Marginal Maximum Likelihood Estimation of Item Parameters: Applications of an EM Algorithm*, Psychometrika, Vol. 46, pp. 443-459.

Chen Y., Wenig C. (2009), *Mining fuzzy association rules from questionnaire data*, Knowledge-Based Systems Journal, Vol. 22, No. 1, pp. 46-56.

Chu H.C., Hwang G. J., Tseng J.C.R., Hwang G. H. (2006), *A computerized approach*

to diagnosing student learning problems in health education, Asian Journal of Health and Information Sciences, Vol. No. 1, pp. 43-60.

Dimauro G., Impedovo S., Modugno R., Pirlo G. (2006), *The e-learning Project at the University of Bari*, WSEAS Transactions on Advances in Engineering Education, Issue 2, Vol. 3, pp. 75-79

Dimauro G., Impedovo S., Pirlo G. (2003), *Traditional learning toword on-line learning*, Proc. International Conference on Technology-enhanced Learning TEL '03, Milano, pp. 355-360.

Fraley R. C., Waller N. G., Brennan K. A. (2000), *An Item Response Theory Analysis of Self-Report Measures of Adult Attachment*, Journal of Personality and Social Psychology, Vol. 78, No. 2, pp. 350-365.

Goldberg D.E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison–Wesley, New York.

Greco N., Dimauro G., Ferrante A., Impedovo S., Lucchese M.G., Modugno R., Pirlo G., Sarcinella L. (2006a), *The PROTEO Project: New Advancements in e-learning activities at the University of Bari*, WSEAS Transactions on Communications, Issue 1, Vol. 5, pp. 23-30.

Greco N., Impedovo D., Pirlo G. (2006b), *New Steps toward the Effective Evaluation of E-learning Activities: A Participant-Based Approach*, WSEAS Transactions on Advances in Engineering Education, Issue 7, Vol. 3, pp. 662-666.

Impedovo D., Pirlo G., Stasolla E. (2011), *Ambienti Virtuali Integrati per Attività Collaborative Real-Time: il prototipo Co.S.M.O.S.*, Journal of e-Learning and Knowledge Society (Je-LKS), Vol 7, No 2, pp. 59-68.

Impedovo S., Lucchese M.G., Pirlo G. (2006), *e-Examinations: an Advanced Methodology for Student's tests on e-Learning University Courses*, WSEAS Transactions on Advances in Engineering Education, Issue 5, Vol. 3, pp. 361- 366

J. W. Strijbos (2011), *Assessment of (Computer-Supported) Collaborative Learning*, IEEE Trans. on Learning Technologies, Vol. 4, No. 1, pp. 59-73.

Kuechler W. L., Simkin M. G. (2003), *How well do multiple choice tests evaluate student understanding in computer programming classes?*, Journal Information System Education, Vol. 14, No. 4, pp. 389–399.

Lan C. H., Graf S., Lai K. R., Kinshuk (2011), *Enrichment of Peer Assessment with Agent Negotiation*, IEEE Trans. on Learning Technologies, Vol. 4, No. 1, pp. 35-46.

Michalewicz Z. (1996), *Genetic Algorithms + Data Structure=Evolution Programs*, Springer Verlag, Berlin, Germany.

Pechenizkiy M., Calders T., Vasilyeva E., De Bra P. (2008), *Mining the student assessment data: lessons drawn from a small scale case study*, International Conference on Educational Data Mining, Cordoba, Spain, pp. 187-191.

Pirlo G. et al. (2008), *Quality enhancement in elearning activities: improvements by mean of a newly engineered e-learning survey*, WSEAS Transactions on Advances in Engineering Education, Vol. 5, No. 4, pp.242-251.

Romero C., Ventura S. (2010), *Educational Data Mining: A Review of the State-of-the-Art*, IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and

Reviews, Vol. 40, No. 6, pp. 601–618.

Romero C., Ventura S., De Bra P. (2009), *Using Mobile and Web-Based Computerized Tests to Evaluate University Students*, Computer Applications in Engineering Education, Vol. 17, No. 4, 435-447.

Romero C., Ventura S., Salcines P. E. (2008), *Data mining in course management systems: Moodle case study and tutorial*, Computers & Education, Vol. 51, No. 1, pp. 368-384.

Van der Linen W.J., Hambleton R.K. (eds.) (1997), *Handbook of modern item response theory*, Springer, New York.

Yamanishi K. and Li H. (2001, *Mining from open answers in questionnaire data*, Proceedings of the Seventh ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, pp. 443-449.