

# Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση

Τόμ. 1 (2011)

2ο Πανελλήνιο Συνέδριο: «Ένταξη και χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία»



2ο Πανελλήνιο Συνέδριο:  
«Ένταξη και χρήση των ΤΠΕ  
στην Εκπαιδευτική Διαδικασία»

Πάτρα

28 - 30 Απριλίου 2011

ISSN : 2529-0924

Εκτίμηση ποιότητας τεστ αυτοαξιολόγησης σε  
Συστήματα Διαχείρισης Μάθησης

Π. Φώταρης, Θ. Μάστορας

## Βιβλιογραφική αναφορά:

Φώταρης Π., & Μάστορας Θ. (2023). Εκτίμηση ποιότητας τεστ αυτοαξιολόγησης σε Συστήματα Διαχείρισης Μάθησης. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 1, 0981–0990. ανακτήθηκε από <https://eproceedings.epublishing.ekt.gr/index.php/cetpe/article/view/4846>

## Εκτίμηση ποιότητας τεστ αυτοαξιολόγησης σε Συστήματα Διαχείρισης Μάθησης

Π. Φώταρης<sup>1</sup>, Θ. Μάστορας<sup>2</sup>

Πανεπιστήμιο Μακεδονίας, Τμήμα Εφαρμοσμένης Πληροφορικής, <sup>1</sup>raf@uom.gr,  
<sup>2</sup>mastoras@uom.gr

### Περίληψη

Με δεδομένο το ότι η ανάπτυξη αξιόπιστου και έγκυρου υλικού αξιολόγησης της μάθησης αποτελεί μια ιδιαίτερα χρονοβόρα και επίπονη διαδικασία, ολοένα και περισσότερα ακαδημαϊκά ιδρύματα χρησιμοποιούν σήμερα εργαλεία Computer Aided Assessment (CAA) για να βελτιώσουν την ποιότητα και την αξιοπιστία των αποτελεσμάτων που προκύπτουν από τα διάφορα τεστ αξιολόγησης της μάθησης. Στην παρούσα εργασία προτείνεται ένα μεθοδολογικό και αρχιτεκτονικό πλαίσιο ενσωμάτωσης ενός εργαλείου CAA σε ένα Σύστημα Διαχείρισης Μάθησης (ΣΔΜ) προκειμένου να επεκταθεί η λειτουργικότητα του τελευταίου με στατιστική ανάλυση βασισμένη στη θεωρία απόκρισης ερωτήματος (IRT). Επίσης περιγράφεται μια εύχρηστη μεθοδολογία ερμηνείας και αξιοποίησης των αποτελεσμάτων της ανάλυσης IRT από τον εκάστοτε δημιουργό μαθησιακού υλικού ανεξαρτήτως του μαθηματικού του υπόβαθρου. Η εφαρμογή των προτεινόμενων μεθοδολογιών αποσκοπεί στη διασφάλιση της υψηλής ποιότητας των τεστ αυτοαξιολόγησης και την αποτελεσματική επαναχρησιμοποίηση των καταλλήλων από τις ερωτήσεις τους.

**Λέξεις-κλειδιά:** Συστήματα Διαχείρισης Μάθησης, αξιολόγηση μάθησης, θεωρία απόκρισης ερωτήματος

### 1. Εισαγωγή

Οι ραγδαίες εξελίξεις στον χώρο των τεχνολογιών πληροφορίας και τηλεπικοινωνιών που προέκυψαν ως αποτέλεσμα της παγκοσμιοποίησης, της διεθνούς ανταγωνιστικότητας και της φιλοσοφίας της ελεύθερης αγοράς, σε συνδυασμό με την ανάγκη για εξοικονόμηση χρόνου και χρήματος καθώς και για παροχή πιο ευέλικτων μορφών εκπαίδευσης, κατέστησαν επιτακτική τη χρήση Συστημάτων Διαχείρισης Μάθησης (Learning Management Systems – LMSs) και εργαλείων αξιολόγησης εξετάσεων χρηστών (Computer Aided Assessment – CAA) στο χώρο των ακαδημαϊκών ιδρυμάτων (Anantatmula & Stankosky, 2008). Τα συγκεκριμένα εργαλεία χρησιμοποιούν συνήθως τεστ αυτοαξιολόγησης με ερωτήσεις πολλαπλής επιλογής ως μέσα αξιολόγησης της μάθησης (Carter et al., 2003). Χάρη σε αυτά είναι δυνατή η παροχή αξιόπιστης εκτίμησης του επιπέδου γνώσεων των εκπαιδευόμενων (Hindi, Najdawi, & Jolo, 2008), ενώ ταυτόχρονα προκαλείται μείωση του φόρτου εργασίας από την πλευρά των εκπαιδευτών λόγω της χορήγησης άμεσης ανατροφοδότησης και αυτόματης βαθμολόγησης (Baggott & Rayne, 2001).

Ωστόσο, επιστημονικές μελέτες καταδεικνύουν ότι ένα ποσοστό που μπορεί να αγγίξει το 50% επί του συνόλου των ερωτήσεων ενός τεστ είναι προβληματικές από το αρχικό στάδιο της συγγραφής τους (Haladyna, 1999), οδηγώντας έτσι σε μη αξιόπιστες εκτιμήσεις των επιδόσεων των εξεταζόμενων (Murphy & Davidhofer, 1988). Συνεπώς, κρίνεται σκόπιμη η ύπαρξη κάποιου μηχανισμού ποιοτικού ελέγχου των τεστ που θα εντοπίζει και θα υποδεικνύει τις ακατάλληλες ερωτήσεις στους εκπαιδευτές προκειμένου αυτοί να τις διορθώσουν ή να τις απομακρύνουν.

Για να εντοπιστούν οι προβληματικές ερωτήσεις που δυσχεραίνουν τη μάθηση είναι απαραίτητη η δυνατότητα εκτίμησης ορισμένων παραμέτρων τους (δυσκολία, διακριτικότητα, τυχαιότητα), καθώς και μίας ή περισσότερων μεταβλητών ενδιαφέροντος, όπως π.χ. η έννοια της γνώσης που έχει αποκτήσει ο εκπαιδευόμενος, ο βαθμός ικανότητας που κατέχει ή η γενικότερη επάρκειά του στο θέμα που διδάσκεται. Τις περισσότερες φορές οι μεταβλητές αυτές δεν μπορούν να παρατηρηθούν άμεσα γιατί είναι λανθάνουσες (latent variables), ή μερικώς παρατηρήσιμες (partial observable). Εντούτοις, μπορούν να μετρηθούν υπολογιστικά εφαρμόζοντας τις αρχές διαφόρων στατιστικών μοντέλων, με επικρατέστερα το μοντέλο ανάλυσης ερωτήσεων (Item Analysis – IA) που βασίζεται στις αρχές της κλασικής θεωρίας μέτρησης (Classical Test Theory – CTT) (Hambleton, 1994; Lord & Novick, 1968), και το μοντέλο της θεωρίας απόκρισης ερωτήματος (Item Response Theory – IRT) (Lord, 1980). Και στις δύο περιπτώσεις αναλύονται αρχικά τα δεδομένα που προέρχονται από τις απαντήσεις (responses) των εξεταζόμενων σε σειρές ερωτήσεων (items) των τεστ. Στη συνέχεια εφαρμόζονται σε αυτά στατιστικές μέθοδοι για την εκτίμηση συγκεκριμένων δεικτών, βάσει των οποίων πραγματοποιείται η τελική αξιολόγηση της καταλληλότητας των ερωτήσεων του τεστ.

## 2. Θεωρία Απόκρισης Ερωτήματος (Item Response Theory)

Μολονότι τα στατιστικά μοντέλα που βασίζονται στη CTT είναι τα επικρατέστερα για την αξιολόγηση τεστ στις κοινωνικές επιστήμες, στην Ψυχομετρία έχουν ξεπεραστεί από τα πιο εξελιγμένα μοντέλα της IRT. Βασική αιτία αποτελεί το γεγονός ότι στη CTT ο υπολογισμός παραμέτρων μιας ερώτησης – π.χ. του δείκτη δυσκολίας – εξαρτάται άμεσα από το δείγμα των εξεταζόμενων (Schmeiser & Welch, 2006). Για παράδειγμα, οι ερωτήσεις ενός τεστ ενδέχεται να φανούν αρκετά εύκολες σε ένα σύνολο ατόμων υψηλού επιπέδου κατάρτισης. Αντιθέτως, αν το ίδιο τεστ δοθεί σε εξεταζόμενους με λιγότερες γνώσεις ή μικρότερες δυνατότητες, ο βαθμός δυσκολίας των ερωτήσεων θα τους φανεί ιδιαίτερα μεγάλος. Με τη χρήση των μοντέλων IRT δεν υπάρχει τέτοιου είδους προκατάληψη (bias) γιατί η δυσκολία μιας ερώτησης καθορίζει απλώς ένα σημείο στο διάστημα των ικανοτήτων του εκάστοτε εξεταζόμενου (Hambleton, 1994).

Η ανάλυση IRT χρησιμοποιεί τρεις εκτιμώμενες παραμέτρους για κάθε ερώτηση ενός τεστ (Yen & Fitzpatrick, 2006):

- Την παράμετρο διακριτικότητας (discrimination parameter),  $a_i$ , η οποία καθορίζει

τη δυνατότητα της ερώτησης να διακρίνει δύο εξεταζόμενους που εμφανίζουν σχετικά μικρή διαφορά στην ικανότητα τους. Αν και θεωρητικά η τιμή της ορίζεται στο διάστημα  $[-\infty, +\infty]$ , στην πράξη οι ερωτήσεις με αρνητική διακριτικότητα απορρίπτονται από τα τεστ αξιολόγησης.

- Την παράμετρο δυσκολίας (difficulty parameter),  $b_i$ , της οποίας η τιμή αυξάνεται όσο μια ερώτηση γίνεται δυσκολότερη.
- Την παράμετρο τυχειότητας (guessing / pseudo-guessing parameter),  $c_i$ , η οποία παριστάνει την πιθανότητα οι εξεταζόμενοι χαμηλής ικανότητας να δώσουν σωστή απάντηση κατά τύχη (π.χ. σε μια ερώτηση πολλαπλής επιλογής ή σωστού – λάθους). Μια αποδεκτή ερώτηση ενός τεστ αξιολόγησης χαρακτηρίζεται σε επίπεδο παραμέτρων IRT από υψηλή διακριτικότητα ( $a > 1$ ), ορθογώνια κατανομή δυσκολίας ( $b$ ) και χαμηλή τυχειότητα ( $c < 0.2$ ) (Baker, 1992). Η πληροφορία που παρέχει η ανάλυση IRT δεν βοηθά μόνο στην αξιολόγηση της επίδοσης ενός εξεταζόμενου αλλά και στη βελτίωση της ποιότητας των ίδιων των ερωτήσεων του τεστ. Έχοντας καθορίσει συγκεκριμένα κριτήρια καταλληλότητας των ερωτήσεων στο προφίλ του εκάστοτε τεστ, οι συγγραφείς του τελευταίου μπορούν να χρησιμοποιήσουν τα παραγόμενα αποτελέσματα για να διακρίνουν αν μια ερώτηση δύναται να επαναχρησιμοποιηθεί αυτούσια, χρειάζεται αναθεώρηση ή είναι ακατάλληλη και πρέπει να απομακρυνθεί από το τεστ.

Οι περισσότεροι εκπαιδευτικοί αποφεύγουν την ανάλυση IRT λόγω της αυξημένης πολυπλοκότητας των υπολογιστικών μεθόδων που απαιτούνται για την εκτίμηση των παραμέτρων της (Lin et al., 2009). Καθώς οι συγκεκριμένες μέθοδοι δεν υποστηρίζονται άμεσα από δημοφιλή στατιστικά πακέτα όπως τα SPSS και SAS, επιβάλλεται η χρήση εξειδικευμένων προγραμμάτων που υστερούν σημαντικά σε φιλικότητα προς τους χρήστες (Yu, 2005), ενώ παράλληλα προϋποθέτουν την ύπαρξη ισχυρού θεωρητικού υποβάθρου στατιστικής για την κατανόηση των αποτελεσμάτων τους (Costagliola & Fucella, 2009). Δυστυχώς αυτό δεν ισχύει για την πλειοψηφία των δημιουργών τεστ αυτοαξιολόγησης και έτσι οι περισσότερες μελέτες που πραγματεύονται τον εντοπισμό προβληματικών ερωτήσεων υιοθετούν μεθόδους που βασίζονται στη CTT (Costagliola, Ferrucci, & Fucella, 2008; Hung et al., 2004), παραβλέποντας τις αδυναμίες αυτής της θεωρίας χάριν ευκολίας στην εφαρμογή της.

Λαμβάνοντας υπόψη τους παραπάνω προβληματισμούς, η παρούσα εργασία προτείνει ένα μεθοδολογικό και αρχιτεκτονικό πλαίσιο για την ενσωμάτωση ενός απλού και εύχρηστου εργαλείου ανάλυσης IRT σε ένα ανοιχτό (open-source) ΣΔΜ, με τέτοιο τρόπο που να είναι εύκολα εφαρμόσιμος σε οποιοδήποτε μοντέρνο σύστημα ηλεκτρονικής μάθησης. Μέσα από το γνώριμο στους καθηγητές περιβάλλον ενός ΣΔΜ, τα αποτελέσματα της ανάλυσης IRT προσφέρονται με τρόπο κατανοητό αποκρύπτοντας άσκοπες τεχνικές λεπτομέρειες. Το μοντέλο IRT που χρησιμοποιείται είναι το λογιστικό μοντέλο τριών παραμέτρων (3PL) (Birnbaum, 1968), με την ικανότητα  $\theta$  να παρουσιάζεται τόσο στη συνήθη κλίμακα -3 έως 3 αλλά και ως κανονικοποιημένο σκορ 0 έως 100. Όσον αφορά τα όρια και τους κανόνες επί των παραμέτρων  $a$ ,  $b$  και  $c$ , αυτά μπορούν να υπολογιστούν αυτόματα από το ΣΔΜ βάσει του ποσοστού των ερωτήσεων που επιλέγεται να αναθεωρηθούν. Τέλος, ορίζεται μια

συγκεκριμένη μεθοδολογία αξιοποίησης – ερμηνείας – αναθεώρησης των αποτελεσμάτων της ανάλυσης IRT. Χάρη στο προτεινόμενο σύστημα, οι εκπαιδευτές μπορούν πλέον εύκολα να εντοπίσουν τις προβληματικές ερωτήσεις ενός τεστ αυτοαξιολόγησης κάνοντας χρήση ανάλυσης IRT εξαιρετικής αξιοπιστίας χωρίς να χρειάζεται να απομνημονεύσουν δεκάδες σελίδες από τεχνικά εγχειρίδια για εντολές και σενάρια.

### 3. Το εργαλείο ανάλυσης IRT ICL

Προκειμένου να πραγματοποιηθεί στατιστική ανάλυση βασισμένη στη θεωρία απόκρισης ερωτήματος επί των αποτελεσμάτων ενός τεστ αυτοαξιολόγησης, είναι απαραίτητη η χρήση εξειδικευμένου λογισμικού (Embretson & Reise, 2000). Αυτή τη στιγμή κυκλοφορούν στην αγορά αρκετά εμπορικά πακέτα που υποστηρίζουν ανάλυση IRT, όπως τα Rascal, Ascal, WINSTEPS, MULTILOG, PARSCALE, RUMM, WINMIRA και BILOG-MG, με το τελευταίο να αποτελεί το δημοφιλέστερο εργαλείο για ανάλυση διχοτομικών μοντέλων IRT παρά το αφιλόξενο περιβάλλον του. Ωστόσο, ο Hanson (2002) ανέπτυξε ένα αυτόνομο (stand-alone) ανοιχτό λογισμικό εκτίμησης των παραμέτρων IRT υπό το όνομα IRT Command Language (ICL), το οποίο παρέχει εξίσου ακριβή και αξιόπιστα αποτελέσματα με το εμπορικό BILOG-MG (Mead, Morris, & Blitz, 2007), ενώ η ανοιχτή φύση του υποστηρίζει κάθε είδους τροποποίηση και επέκταση. Στην πραγματικότητα το ICL αποτελείται από ένα σύνολο συναρτήσεων εκτίμησης των παραμέτρων IRT (ETIRM) (Hanson, 2000) ενσωματωμένων στη γλώσσα προγραμματισμού TCL (tickle) (Welch, Jones, & Hobbs, 2003), πράγμα που του επιτρέπει να πραγματοποιεί ιδιαίτερα πολύπλοκους υπολογισμούς. Επιπλέον, το ICL μπορεί να εκτελείται στο παρασκήνιο μέσω της γραμμής εντολών, καθώς και να εξάγει τα αποτελέσματα της ανάλυσης IRT σε αρχεία κειμένου. Βάσει αυτών των χαρακτηριστικών και με δεδομένο το ότι το προτεινόμενο πλαίσιο χρησιμοποιεί μόνο ένα διχοτομικό μοντέλο IRT τριών παραμέτρων (3PL), το ICL αποδείχθηκε παραπάνω από ικανοποιητικό, γι' αυτό και επιλέχθηκε να συνοδεύσει το ΣΔΜ στην αξιολόγηση των τεστ.

### 4. Ενσωμάτωση ανάλυσης IRT στο ΣΔΜ Dokeos

Το Dokeos είναι ένα ανοιχτό ΣΔΜ που διανέμεται δωρεάν με τη Γενική Άδεια Δημόσιας Χρήσης (General Public License) του Ιδρύματος Ελεύθερου Λογισμικού (Free Software Foundation). Έχει υλοποιηθεί στη γλώσσα διαδικτυακού προγραμματισμού PHP και απαιτεί τον διακομιστή Παγκόσμιου Ιστού Apache και το Σύστημα Διαχείρισης Βάσης Δεδομένων MySQL για να μπορέσει να λειτουργήσει. Εδώ και τέσσερα χρόνια το Dokeos χρησιμοποιείται επιτυχώς σε δύο ακαδημαϊκά μαθήματα του Πανεπιστημίου Μακεδονίας, λαμβάνοντας ικανοποιητικές κριτικές τόσο από τους εκπαιδευτές όσο και από τους φοιτητές. Για να μπορέσει να επεκταθεί η λειτουργία του με ανάλυση IRT χρειάστηκε να τροποποιηθεί ο υπάρχων πηγαίος κώδικας. Χάρη σε αυτές τις αλλαγές το εν λόγω ΣΔΜ μπορεί πλέον να υποστηρίξει μια σειρά από νέες λειτουργίες. Πιο συγκεκριμένα:



**Σχήμα 1:** Λειτουργίες που υποστηρίζει η διευρυμένη έκδοση του ΣΔΜ Dokeos

1. Μετά το πέρας ενός τεστ αυτοαξιολόγησης, το ΣΔΜ έχει τη δυνατότητα να αποθηκεύσει στη βάση δεδομένων του τις απαντήσεις των εξεταζόμενων για κάθε ερώτηση του τεστ, αντί να διατηρεί μόνο ένα τελικό σκορ.
2. Ο συγγραφέας του τεστ μπορεί να καθορίσει τα επιτρεπτά όρια για τον δείκτη διακριτικότητας και τον βαθμό δυσκολίας των ερωτήσεων, καθώς και για την παράμετρο τυχασιότητας. Στην πράξη ορίζει: α) κάτω όριο του δείκτη διακριτικότητας, β) άνω και κάτω όριο του βαθμού δυσκολίας και γ) άνω όριο της παραμέτρου τυχασιότητας.
3. Το ΣΔΜ αποθηκεύει τις προαναφερθείσες τιμές ως κανόνες εγκυρότητας στο προφίλ του εκάστοτε τεστ, ενώ υπάρχει η δυνατότητα αυτόματης συμπλήρωσής τους όταν ζητείται να αποκλειστεί ένα συγκεκριμένο ποσοστό ερωτήσεων. Για παράδειγμα, είναι δυνατό με μια απλή ταξινόμηση των υπολογισμένων δεικτών διακριτικότητας να συμπληρωθεί αυτόματα η τιμή κάτω από την οποία βρίσκεται το 10% των ερωτήσεων (Σχήμα 1.1).
4. Κάθε φορά που ο συγγραφέας του τεστ ζητά εκτέλεση της ανάλυσης IRT, μεταβαίνει σε μια σελίδα όπου παρουσιάζονται οι εκτιμώμενες παράμετροι δυσκολίας, διακριτικότητας και τυχασιότητας για κάθε ερώτηση του τεστ. Αν αυτές οι παράμετροι παραβιάζουν κάποια από τα όρια που ήδη καταχωρήθηκαν ως κανόνες εγκυρότητας στο προφίλ του συγκεκριμένου τεστ, τότε οι αντίστοιχες ερωτήσεις

επισημαίνονται προκειμένου να αναθεωρηθούν αργότερα (Σχήμα 1.2). Αφού ολοκληρωθεί ο έλεγχος εγκυρότητας των παραμέτρων IRT όλων των ερωτήσεων, ο συγγραφέας του τεστ μπορεί να χρησιμοποιήσει τα παραγόμενα αποτελέσματα για να αποφασίσει ποιες από τις ερωτήσεις χρειάζονται απομάκρυνση, τροποποίηση ή είναι κατάλληλες προς χρήση.

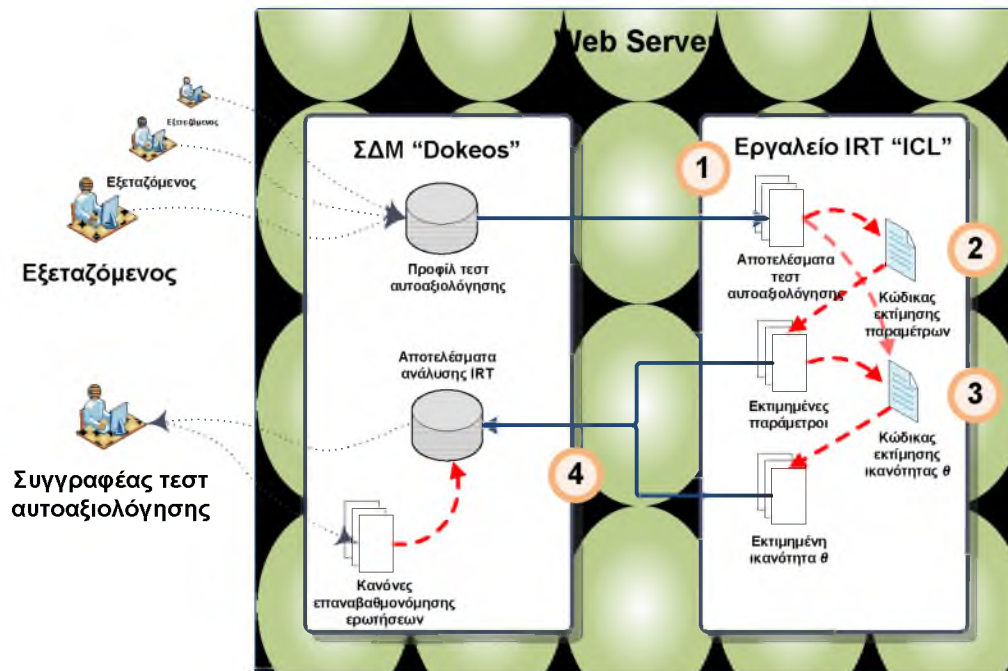
5. Όταν ζητηθούν τα τελικά αποτελέσματα ενός τεστ αυτοαξιολόγησης εκτελείται αυτόματα ανάλυση IRT έτσι ώστε το τελικό σκορ του κάθε εξεταζόμενου να εμφανιστεί τόσο με την κλασική του μορφή όσο και με τη μορφή εκτίμησης της ικανότητάς του  $\theta$ , όπως αυτή προέκυψε από τα αποτελέσματα της ανάλυσης IRT. Για λόγους ευκολότερης ερμηνείας, το  $\theta$  παρουσιάζεται τόσο στη συνήθη κλίμακα -3 έως 3 όσο και ως κανονικοποιημένο σκορ 0-100 (Σχήμα 1.3).

## 5. Μεθοδολογία ανάλυσης ερωτήσεων

Η προτεινόμενη μεθοδολογία ανάλυσης ερωτήσεων αποτελείται από τέσσερα βήματα, με καθένα από αυτά να αποτελεί εκτελούμενη ενέργεια του ΣΔΜ. Επιπλέον, το αρχικό σχήμα (schema) της βάσης δεδομένων έχει επεκταθεί έτσι ώστε να υποστηρίζει ορισμένες επιπρόσθετες λειτουργίες. Αν και το ΣΔΜ που χρησιμοποιείται στην παρούσα μελέτη είναι το Dokeos, αυτό δεν αποκλείει την εφαρμογή της συγκεκριμένης μεθοδολογίας ανάλυσης ερωτήσεων σε διαφορετικά εργαλεία ηλεκτρονικής μάθησης. Η ανάλυση IRT ελέγχεται από τον κώδικα PHP του ΣΔΜ. Κάθε φορά που απαιτείται ενημέρωση των αποτελεσμάτων της ανάλυσης IRT, το ΣΔΜ εξάγει τα κατάλληλα αρχεία δεδομένων (data) και κώδικα (script) TCL. Στη συνέχεια το ΣΔΜ εκτελεί έναν αριθμό από κλήσεις στο πρόγραμμα ICL και αφού αναλύσει συντακτικά (parsing) τα αποτελέσματα της ανάλυσης IRT, τα εισάγει στη βάση δεδομένων του. Στη συνέχεια δίνεται μια λεπτομερέστερη περιγραφή των τεσσάρων βημάτων της εν λόγω μεθοδολογίας:

- **Βήμα 1:** Το ΣΔΜ εξάγει τα αποτελέσματα του τεστ (δηλαδή την βαθμολογία) σε ένα αρχείο δεδομένων και προετοιμάζει κατάλληλα το αρχείο κώδικα TCL που πρόκειται να τα επεξεργαστεί (αρχείο κώδικα εκτίμησης παραμέτρων).
- **Βήμα 2:** Το ΣΔΜ καλεί και εκτελεί το ICL δίνοντάς του ως παράμετρο το παραπάνω αρχείο κώδικα εκτίμησης παραμέτρων. Αυτό έχει ως αποτέλεσμα τη δημιουργία στον δίσκο ενός αρχείου δεδομένων που περιέχει τις εκτιμώμενες χαρακτηριστικές παραμέτρους  $a$ ,  $b$  και  $c$  για κάθε ερώτηση. Ταυτόχρονα το ΣΔΜ ετοιμάζει ένα ακόμη αρχείο κώδικα TCL, το οποίο θα επεξεργαστεί αυτές τις παραμέτρους ώστε να εκτιμηθεί και το  $\theta$  (αρχείο κώδικα εκτίμησης  $\theta$ ).
- **Βήμα 3:** Το ΣΔΜ καλεί και εκτελεί το ICL δίνοντας του ως παράμετρο το παραπάνω αρχείο κώδικα εκτίμησης  $\theta$ , έτσι ώστε να δημιουργηθεί στον δίσκο ένα αρχείο δεδομένων με τις αξιολογήσεις – δηλ. την ικανότητα  $\theta$  – των εξεταζόμενων.
- **Βήμα 4:** Τέλος, το ΣΔΜ διαβάζει από τον δίσκο τα δύο αρχεία αποτελεσμάτων του ICL (\*.par και \*.theta), τα οποία και εισάγει στη βάση δεδομένων του προκειμένου να επεκτείνει τη λειτουργία του με υπηρεσίες ανάλυσης IRT και επαναβαθμονόμησης των παραμέτρων των ερωτήσεων του τεστ (item calibration). Η

αρχιτεκτονική του συστήματος που ακολουθεί τη συγκεκριμένη προσέγγιση απεικονίζεται στο Σχήμα 2.



Σχήμα 2: Αρχιτεκτονική προτεινόμενου συστήματος

## 6. Μεθοδολογία αξιοποίησης και ερμηνείας αποτελεσμάτων

Προκειμένου να επιτευχθεί η βελτίωση των προβληματικών ερωτήσεων του τεστ κρίνεται σκόπιμος ο καθορισμός και η υιοθέτηση μιας συγκεκριμένης μεθοδολογίας αξιοποίησης και ερμηνείας των αποτελεσμάτων της ανάλυσης IRT. Γι' αυτόν τον λόγο προτείνεται ο εξής κύκλος ανάλυσης – ερμηνείας – αναθεώρησης (Σχήμα 3):

- **Βήμα 1:** Αρχικά ο εκπαιδευτής / συγγραφέας του τεστ αυτοαξιολόγησης εισάγει στο ΣΔΜ τους επιθυμητούς κανόνες εγκυρότητας των ερωτήσεων (κάτω όριο του δείκτη διακριτικότητας, άνω και κάτω όριο του βαθμού δυσκολίας, άνω όριο της παραμέτρου τυχαιότητας) για το εκάστοτε τεστ.
- **Βήμα 2:** Στη συνέχεια πραγματοποιείται η διεξαγωγή του τεστ.
- **Βήμα 3:** Ακολουθεί εφαρμογή της ανάλυσης IRT επί των αποτελεσμάτων του τεστ από το ΣΔΜ. Αν οι εκτιμώμενες τιμές των παραμέτρων IRT μιας ερώτησης παραβιάζουν τους κανόνες εγκυρότητας που ανήκουν στο προφίλ του συγκεκριμένου τεστ, τότε η συγκεκριμένη ερώτηση επισημαίνεται.
- **Βήμα 4:** Όταν έχουν εντοπιστεί όλες οι προβληματικές ερωτήσεις από το ΣΔΜ, ο συγγραφέας του τεστ τις μελετά προκειμένου να ερμηνεύσει τα βαθύτερα αίτια του προβλήματος σε κάθε περίπτωση (π.χ. ασαφείς ή διαφορούμενες εκφράσεις, τυπογραφικά ή λογικά λάθη, έλλειψη ουσιαστικών πληροφοριών κ.λπ.).

- **Βήμα 5:** Με βάση την παραπάνω ερμηνεία των αποτελεσμάτων ο συγγραφέας του τεστ προβαίνει στις κατάλληλες τροποποιήσεις των ερωτήσεων, δημιουργώντας έτσι μια καινούργια έκδοση του τεστ στην οποία οι ελαττωματικές ερωτήσεις έχουν διορθωθεί ή αντικατασταθεί. Μετά την ολοκλήρωση των αναθεωρήσεων, το νέο τεστ είναι πλέον έτοιμο προς χρήση. Αφού διεξαχθεί ένας ακόμα κύκλος εξετάσεων, μια νέα έκθεση ανάλυσης θα επιβεβαιώσει εάν όλες οι ερωτήσεις συμφωνούν με τους κανόνες εγκυρότητας.



**Σχήμα 3:** Κύκλος ανάλυσης – ερμηνείας – αναθεώρησης

Ο αριθμός των επαναλήψεων της αναθεώρησης ενός συγκεκριμένου τεστ, ο οποίος είναι απαραίτητος για την εξάλειψη όλων των προβληματικών ερωτήσεων, εξαρτάται από την κατανόηση των αποτελεσμάτων της ανάλυσης. Όσο γρηγορότερα εντοπίσει ο συγγραφέας του τεστ την πραγματική αιτία κάθε προβλήματος και βρει την κατάλληλη λύση, τόσο λιγότερες επαναλήψεις χρειάζονται.

## 7. Συμπεράσματα

Η παρούσα εργασία παρουσίασε ένα μεθοδολογικό και αρχιτεκτονικό πλαίσιο για την επέκταση ενός ΣΔΜ με ανάλυση IRT επί των αποτελεσμάτων των τεστ αυτοαξιολόγησης. Χάρη στην κατάλληλη αξιοποίηση του εργαλείου ανάλυσης ICL, το προτεινόμενο πλαίσιο απαλλάσσει τον προγραμματιστή από την υλοποίηση των δύσκολων αλγορίθμων ανάλυσης IRT, ενώ ταυτόχρονα εγγυάται αξιόπιστα αποτελέσματα, τα οποία ενσωματώνονται στο γνώσιμο περιβάλλον ενός ΣΔΜ. Χρησιμοποιώντας συγκεκριμένους κανόνες εγκυρότητας, το σύστημα μπορεί να

αναγνωρίσει τις προβληματικές ερωτήσεις του τεστ και να τις εμφανίσει στον εκάστοτε εκπαιδευτή προς αναθεώρηση. Μελλοντικό στόχο της παρούσας ερευνητικής εργασίας αποτελεί η μελέτη της ακρίβειας του εντοπισμού αυτών των ερωτήσεων σε μικρά πλήθη εξεταζόμενων, τα οποία αποτελούν και τη συνηθέστερη περίπτωση σε ακαδημαϊκά περιβάλλοντα.

### **Βιβλιογραφία**

- Anantatmula, V. S., & Stankosky, M. (2008). KM criteria for different types of organisations. *International Journal of Knowledge and Learning*, 4(1), 18-35.
- Baggott, G., & Rayne, R. (2001). Learning Support for Mature, Part-time, Evening Students: Providing Feedback via Frequent, Computer-Based Assessments. *Fifth International Computer Assisted Assessment Conference*, Loughborough University.
- Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker.
- Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring an Examinee's Ability. Στο F. M. Lord and M. R. Novick (επιμ.), *Statistical Theories of Mental Test Scores*, Reading, UK: Addison-Wesley.
- Carter, J., Ala-Mutka, K., Fuller, U., Dick, M., English, J., Fone, W., & Sheard, J. (2003). How shall we assess this? *ACM SIGCSE Bulletin, Working group reports from ITiCSE on Innovation and technology in computer science education*, 35(4), 107-123.
- Costagliola, G., Ferrucci, F., & Fucella, V. (2008). *A Web-Based E-Testing System Supporting Test Quality Improvement*. Proceedings of the Advances in Web Based Learning – ICWL 2007, Edinburgh, UK.
- Costagliola, G., & Fucella, V. (2009). A Rule-Based E-Testing System for Test Quality Improvement. *International Journal of Distance Education Technologies*, 7(2), 63-82.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Haladyna, T. M. (1999). *Developing and Validating Multiple-Choice Test Items*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hambleton, R. K. (1994). Item Response Theory: A Broad Psychometric Framework for Measurement Advances. *Psicothema*, 6(3), 535-556.
- Hanson, B. A. (2000). *Estimation Toolkit for Item Response Models (ETIRM)*. Ανακτήθηκε 18/3/2010, από τη διεύθυνση <http://www.b-a-h.com/software/cpp/etirm.html>.

- Hanson, B. A. (2002). *IRT Command Language (ICL)*. Ανακτήθηκε 18/3/2010, από τη διεύθυνση <http://www.b-a-h.com/software/irt/icl/index.html>.
- Hindi, N. M., Najdawi, M. K., & Jolo, H. A. M. (2008). An Examination of Assessment Practices in Colleges of Business at Various Middle East Countries. *International Journal of Teaching and Case Studies*, 1(4), 319-332.
- Hung, J.-C., Lin, L.-J., Chang, W., Shih, T.-K., Hsu, H., Chang, H.-B., Chang, H.-P., & Huang, K. (2004). *A Cognition Assessment Authoring System for E-Learning*. Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCS 2004 Workshops).
- Lin, X., Chen, H., Mather, R., & Fletcher, H. (2009). *Adaptive Assessment – A Practice of Classification on Small-Size Training Sets*. Proceedings of the Society for Information Technology & Teacher Education International Conference (SITE 2009), Chesapeake, VA.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.
- Mead, A. D., Morris, S. B., & Blitz, D. L. (2007). *Open-source IRT: A Comparison of BILOG-MG and ICL Features and Item Parameter Recovery*. Illinois Institute of Technology, Institute of Psychology, Chicago, IL. Ανακτήθηκε 18/3/2010, από τη διεύθυνση <http://mypages.iit.edu/~mead/MeadMorrisBlitz2007.pdf>
- Murphy, K. R., & Davidhofer, C. O. (1988). *Psychological Testing Principles and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Paxton, M. (2001). A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education*, 25(2), 109-119.
- Schmeiser, C. B., & Welch, C. J. (2006). Test Development. Στο R. L. Brennan (επιμ.), *Educational Measurement* (4th ed.). Westport, CT: Praeger Publishers.
- Welch, B. B., Jones, K., & Hobbs, J. (2003). *Practical programming in Tcl and Tk* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item Response Theory. Στο R. L. Brennan (επιμ.), *Educational Measurement* (4th ed.). Westport, CT: Praeger Publishers.
- Yu, C.-H. (2005). A Simple Guide to the Item Response Theory (IRT). Ανακτήθηκε 18/3/2010, από τη διεύθυνση <http://seamonkey.ed.asu.edu/~alex/computer/sas/IRT.pdf>