

## International Conference in Open and Distance Learning

Vol 9, No 2A (2017)

Ο Σχεδιασμός της Μάθησης

**Τόμος 2, Μέρος Α**

### Πρακτικά

9<sup>ο</sup> Διεθνές Συνέδριο για την Ανοικτή  
& εξ Αποστάσεως Εκπαίδευση

Αθήνα, 23 – 26 Νοεμβρίου 2017

### Ο Σχεδιασμός της Μάθησης

Επιμέλεια  
Αντώνης Λιοναράκης  
Σύλβη Ιωακειμίδου  
Μαρία Νιάρη  
Γκέλη Μανούσου  
Τόνια Χαρτοφύλακα  
Σοφία Παπαδημητρίου  
Άννα Αποστολίδου

ISBN 978-618-82258-7-9  
ISBN SET 978-618-82258-5-5



Ελληνικό Ανοικτό Πανεπιστήμιο  
Ελληνικό Δίκτυο Ανοικτής & εξ Αποστάσεως Εκπαίδευσης

**Process framework for objectification of selection  
of items for the final assessment through utilization  
of interim assessment data**

*Ιωάννης Σπυρίδωνος Κάτσενος, Γεώργιος Σίλα  
Ανδρουλάκης*

doi: [10.12681/icodl.1044](https://doi.org/10.12681/icodl.1044)

**Πλαίσιο διαδικασίας για την αντικειμενοποίηση της επιλογής ερωτημάτων  
τελικής δοκιμασίας μέσω αξιοποίησης των δεδομένων από ενδιάμεσες  
δοκιμασίες αξιολόγησης**

**Process framework for objectification of selection of items for the final  
assessment through utilization of interim assessment data**

**Ιωάννης Σ. Κάτσης**  
Υποψ. Διδάκτορας  
Τμήμα Διοίκησης Επιχειρήσεων  
Πανεπιστήμιο Πατρών  
[ikatsenos@gmail.com](mailto:ikatsenos@gmail.com)

**Γεώργιος Σ. Ανδρουλάκης**  
Αναπληρωτής Καθηγητής  
Τμήμα Διοίκησης Επιχειρήσεων  
Πανεπιστήμιο Πατρών  
[gandroul@upatras.gr](mailto:gandroul@upatras.gr)

**Abstract**

Choosing the right items for the final assessment test of an educational module as part of an educational program is a demanding task, considering the population to be examined, the educational goals to be evaluated and the comparability of the results over different groups of examinees in different cycles of the educational program.

In this paper, we present a process framework for analyzing interim assessments and selecting the right items for the final assessment of an educational module. The interim assessment tests are analyzed using Item Response Theory and the Items' Information Function (IIF) are determined. The criteria for selecting items are their peak information values, their distribution over the ability scale  $\theta$  and their coverage of the educational goals set. The re-used items are transformed to pseudo-open ended questions and graded according to educational goals achievement. Application of the process framework on two university modules showed that the overall information of the tests is increased from the interim tests to the final test while the average student grades remained at the same levels.

**Keywords:** *Education, Lifelong learning, Learning assessment,*

**Περίληψη**

Η επιλογή των κατάλληλων ερωτημάτων για την τελική δοκιμασία ενός γνωστικού αντικείμενου σε ένα εκπαιδευτικό πρόγραμμα είναι μια απαιτητική διαδικασία, δεδομένων των στοιχείων του πληθυσμού των εξεταζόμενων, των εκπαιδευτικών στόχων που θα πρέπει να αξιολογηθούν και της επιζητούμενης συγκρισιμότητας των αποτελεσμάτων ανάμεσα σε διαφορετικές ομάδες εξεταζόμενων σε διαφορετικούς κύκλους σπουδών των προγράμματος. Στην εργασία αυτή, παρουσιάζουμε ένα πλαίσιο διαδικασίας για την ανάλυση ενδιάμεσων αξιολογήσεων και την επιλογή από αυτές των κατάλληλων ερωτημάτων για την τελική αξιολόγηση ενός γνωστικού αντικείμενου. Οι ενδιάμεσες δοκιμασίες αναλύονται με τη θεωρία απόκρισης ερωτήματος (IRT) και προσδιορίζονται οι συναρτήσεις πληροφορίας ερωτήματος. Τα κριτήρια για την επιλογή ερωτημάτων είναι οι μέγιστες τιμές των συναρτήσεων πληροφορίας, η κατανομή τους στην κλίμακα ικανότητας  $\theta$  και η κάλυψη του εύρους των εκπαιδευτικών στόχων. Τα επαναχρησιμοποιούμενα ερωτήματα, μετατρέπονται σε ψευδο-ανοικτού τύπου και βαθμολογούνται σύμφωνα με το επίπεδο επίτευξης των

αντίστοιχων σε αυτά γνωστικών στόχων. Η εφαρμογή του προτεινόμενου πλαισίου σε δύο πανεπιστημιακά μαθήματα έδειξε ότι η ολική πληροφορία των δοκιμασιών αυξήθηκε από τις ενδιάμεσες στις τελικές δοκιμασίες ενώ η μέση βαθμολογία των φοιτητών παρέμεινε στα ίδια επίπεδα.

**Λέξεις-κλειδιά:** Εκπαίδευση, Δια βίου μάθηση, Εκπαιδευτική αξιολόγηση

## 1. Εισαγωγή

Βασικό μέρος οποιουδήποτε εκπαιδευτικού προγράμματος είναι η αξιολόγηση των συμμετεχόντων σε αυτό μέσω της αξιολόγησης των εκπαιδευόμενων στα επιμέρους γνωστικά αντικείμενα (modules), από τα οποία αυτό δομείται (μαθήματα, θεματικές ενότητες κλπ). Ανάλογα με τη βαθμίδα εκπαίδευσης, το γνωστικό αντικείμενο και τη διδακτική προσέγγιση μπορούν να χρησιμοποιηθούν διάφοροι συνδυασμοί από ένα πλήθος τεχνικών και μέσων αξιολόγησης, όπως για παράδειγμα: ενδιάμεσες δοκιμασίες (τεστ και διαγωνίσματα), εργασίες, εργαστηριακές ασκήσεις, ερωτηματολόγια, παρουσιάσεις σε φυσική ή ηλεκτρονική μορφή για να αξιολογηθεί είτε η πορεία της μάθησης (διαμορφωτική αξιολόγηση), είτε το τελικό αποτέλεσμα (τελική αξιολόγηση), συνήθως, ως προς την επίτευξη προκαθορισμένων εκπαιδευτικών στόχων. Η στροφή από τη δασκαλοκεντρική προσέγγιση προς το μοντέλο των γνωστικών αποτελεσμάτων οδηγεί αφενός στην εκ των προτέρων διατύπωση των αναμενόμενων ικανοτήτων και γνώσεων των εξεταζόμενων στο τέλος της εκπαιδευτικής διαδικασίας, αφετέρου στη διατύπωση των στρατηγικών και των μέσων, που θα τους οδηγήσουν στο κατάλληλο επίπεδο, αλλά και στην εύρεση της κατάλληλης διαδικασίας για τη μέτρηση αυτών των γνωστικών αποτελεσμάτων (Tam, 2014).

Μέρος της διαδικασίας μέτρησης των γνωστικών αποτελεσμάτων είναι τόσο η εκτίμηση της δυσκολίας και της καταλληλότητας των δοκιμασιών για το συγκεκριμένο πληθυσμό-στόχο, όσο και η αντικειμενοποίηση της βαθμολογίας προκειμένου να ελαχιστοποιηθεί η επίδραση παραγόντων, όπως: η τυχαιότητα των απαντήσεων ή η μη-επιθυμητή συνεργασία των εξεταζόμενων. Το τελευταίο μπορεί να βοηθήσει, εφόσον επιτευχθεί σε ικανοποιητικό βαθμό στη **διαχρονική συγκρισιμότητα των αποτελεσμάτων**, ανάμεσα σε διαφορετικές περιόδους του ίδιου εκπαιδευτικού προγράμματος. Η τελική μέτρηση των γνωστικών αποτελεσμάτων υπόκειται ταυτόχρονα σε περιορισμούς χρόνου και διαθέσιμων πόρων, άρα είναι σημαντικό να επιλέγονται οι κατάλληλες δοκιμασίες, που θα δώσουν τα πιο αντιπροσωπευτικά αποτελέσματα.

Σε αυτή την εργασία, παρουσιάζουμε πλαίσιο διαδικασίας επιλογής ερωτημάτων για την τελική αξιολόγηση συμμετεχόντων σε ένα εκπαιδευτικό πρόγραμμα ως προς την επίτευξη προκαθορισμένων εκπαιδευτικών στόχων. Στο προτεινόμενο πλαίσιο διαδικασίας συνδυάζονται δεδομένα των ενδιάμεσων αξιολογήσεων για να οδηγήσουν στην επιλογή των κατάλληλων ερωτημάτων, που συνθέτουν τη δοκιμασία της τελικής αξιολόγησης. Τα στοιχεία, που αποτελούν τα κύρια κριτήρια επιλογής των ερωτημάτων, είναι η μορφή της συνάρτησης πληροφορίας τους και η κατανομή τους στην κλίμακα ικανότητας των εξεταζόμενων.

## 2. Θεωρητικό Πλαίσιο

### 2.1. Εκπαιδευτικοί στόχοι - ταξινομίες εκπαιδευτικών στόχων

Η αξιολόγηση των συμμετεχόντων σε ένα γνωστικό αντικείμενο προϋποθέτει την διατύπωση συγκεκριμένων εκπαιδευτικών στόχων, διότι έτσι: (α) Αυξάνεται η

πιθανότητα οι διδάσκοντες να παρέχουν τις ευκαιρίες στους εκπαιδευόμενους να μαθαίνουν και να εξασκούν τις ζητούμενες γνώσεις και δεξιότητες, που απαιτούνται για την ίδια την αξιολόγηση και (β) καλά αποτελέσματα στην αξιολόγηση είναι περισσότερο πιθανό να σημαίνουν τελικά και καλό επίπεδο γνώσεων ("Align Assessments with Objectives", 2017).

Για τη διατύπωση εκπαιδευτικών στόχων έχουν προταθεί αρκετές ταξινομίες. Η ταξινομία SOLO αναπτύχθηκε από τους Biggs και Collins (Biggs & Collins, 1989), σαν μια μέθοδος περιγραφής της αύξησης του επιπέδου πολυπλοκότητας στην κατανόηση εννοιών. Τα πέντε επίπεδα της ταξινομίας SOLO περιγράφουν την ικανότητα του εκπαιδευόμενου να συνδέσει έννοιες και να συσχετίσει αρχές με νέες ιδέες. Επιπλέον, η πιο πρόσφατα διατυπωμένη ταξινομία του Fink από το 2003 (Fink, 2003) δεν είναι ιεραρχική, αλλά καλύπτει την τομή έξι τομέων σημαντικών για τις αλλαγές που συνεπάγονται μάθηση. Στα ελληνικά πανεπιστήμια, έτσι και στα μαθήματα του Τμήματος Διοίκησης Επιχειρήσεων, που αποτέλεσαν το πεδίο της έρευνάς μας, η αξιολόγηση επιτελείται μόνο στον γνωστικό τομέα, ενώ για παράδειγμα οι τομείς της Ανθρώπινης Διάστασης (Human Dimension), των συναισθημάτων (Caring) και της μάθησης πώς να μαθαίνουμε (Learning how to Learn) της ταξινομίας Fink δεν αντιμετωπίζονται.

Όμως, ο πλέον γνωστός τρόπος για τη διατύπωση εκπαιδευτικών στόχων παραμένει η ταξινομία Bloom (Bloom, 1956) που διατυπώθηκε τη δεκαετία του 1950 και έχει αναθεωρηθεί σε διάφορα σημεία (Anderson et al., 2001). Σύμφωνα με αυτή, οι εκπαιδευτικοί στόχοι ενός μαθήματος εντάσσονται σε τρεις τομείς: γνωστικό, συναισθηματικό και ψυχοκινητικό και ιεραρχούνται σε επίπεδα. Τόσο στην αρχική, όσο και στην αναθεωρημένη ταξινομία Bloom για τη διατύπωση των εκπαιδευτικών στόχων, που είναι συγκεκριμένοι και προσδιορίζουν τους εκάστοτε εκπαιδευτικούς σκοπούς προτείνονται μια σειρά από ρήματα (Anderson et al., 2001). Έτσι, ενδεικτικά, για τη διατύπωση στόχων πρώτου επιπέδου (remember) προτείνονται ρήματα όπως: Αναγνωρίζω, αναφέρω, ορίζω κ.ά., για το τρίτο επίπεδο (apply) εφαρμόζω, χρησιμοποιώ κ.ά., ενώ για το έκτο επίπεδο (create) ρήματα όπως τροποποιώ, συνδυάζω, βελτιώνω, επινοώ. Όμοια και στις άλλες ταξινομίες προτείνονται αντίστοιχα ρήματα για τα διαφορετικά επίπεδα ή τις διαστάσεις που προβλέπουν.

## 2.2. Διαμορφωτική, ενδιάμεση και τελική αξιολόγηση

Οι διάφοροι τύποι εκπαιδευτικής αξιολόγησης έχουν διαφορετικούς σκοπούς, κάποιιοι από τους οποίους είναι μεταξύ τους αντικρουόμενοι. Ενώ τη δεκαετία του 1990 οι κύριοι σκοποί ήταν η υποστήριξη της μάθησης (formative) και η υποβοήθηση της προόδου και η πιστοποίηση (summative), σήμερα αμφισβητείται η διάκριση μεταξύ της διαμορφωτικής (formative) και της αθροιστικής ή ανακεφαλαιωτικής (summative) αξιολόγησης με την έννοια ότι τα αθροιστικά αποτελέσματα μπορούν να χρησιμοποιηθούν ποικιλοτρόπως, άρα και διαμορφωτικά. Επιπλέον, οι όροι "ενδιάμεση" και "τελική" αξιολόγηση σχετίζονται με το σκοπό της αξιολόγησης και τη μετέπειτα χρήση της πληροφορίας, που συγκεντρώνεται ως προς την πρόοδο των εξεταζόμενων ή για μια κρίση της ικανότητας τους στο εξεταζόμενο αντικείμενο και όχι με τη μέθοδο συλλογής των δεδομένων (Isaacs, Zara, Herbert, Coombs., & Smith, 2013).

Έτσι, μια ενδιάμεση αξιολόγηση στο τέλος μιας ενότητας του γνωστικού αντικείμενου μπορεί να χρησιμοποιηθεί, ανάμεσα στα άλλα, για να αξιολογηθεί ο βαθμός επίτευξης των εκπαιδευτικών στόχων μέχρι το δεδομένο χρονικό σημείο, να προβλεφθεί η απόδοση του εξεταζόμενου στην τελική αξιολόγηση, να καθοριστεί αν

μια εκπαιδευτική προσέγγιση είναι καταλληλότερη από μια άλλη, να συγκεντρωθεί αθροίζομενη πληροφορία για τα επιτεύγματα των εξεταζόμενων και να δοθεί στον διδάσκοντα πληροφορία για την αντιμετώπιση των επόμενων ομάδων εκπαιδευόμενων στην ίδια ενότητα κλπ, ενώ η τελική αξιολόγηση κλείνει ένα εκπαιδευτικό πρόγραμμα με σκοπό την συνολική αξιολόγηση των μαθησιακών αποτελεσμάτων των εκπαιδευόμενων (Perie, Marion, & Gong, 2009).

### 2.3. Αποτίμηση αποτελεσμάτων δοκιμασίας - Θεωρία απόκρισης ερωτήματος (Item Response Theory, IRT)

Για την αξιολόγηση των συμμετεχόντων ενός εκπαιδευτικού προγράμματος σε ένα γνωστικό αντικείμενο, μέσω δοκιμασιών οποιασδήποτε μορφής, και την εξαγωγή ατομικών και ομαδικών δεικτών απόδοσης θα μπορούσε να χρησιμοποιηθεί είτε η κλασική θεωρία δοκιμασίας (Classical Test Theory, CTT), είτε η νεότερη θεωρία απόκρισης ερωτήματος (Item Response Theory, IRT). Αν και όχι χωρίς αντίλογο (Fan, 1998) σχετικά με την υπεροχή της, η Θεωρία Απόκρισης Ερωτήματος (IRT) τείνει να επικρατήσει στην αποτίμηση των αποτελεσμάτων των δοκιμασιών αξιολόγησης εκπαιδευτικών αποτελεσμάτων και όχι μόνο (π.χ. εκτίμηση του πόνου ως λανθάνουσα μεταβλητή (Kean & Reilly, 2014)).

Είναι γνωστό ότι ενώ η CTT βασίζεται και αναλύει μια πλήρη δοκιμασία (διαγώνισμα, τεστ), που αποτελείται από πολλά ερωτήματα, η IRT αναλύει κάθε ερώτημα ξεχωριστά, επιτρέποντας στον αναλυτή να προσδιορίσει την πιθανότητα να απαντηθεί ένα οποιοδήποτε ερώτημα μιας δοκιμασίας από τους εξεταζόμενους διαφορετικού επιπέδου ικανότητας  $\theta$ , ως συνάρτηση αυτής της ικανότητας  $\theta$  και κάποιων παραμέτρων του ερωτήματος (π.χ. διακριτική ικανότητα και δυσκολία για το μοντέλο δύο παραμέτρων).

Το γενικό τετραπαραμετρικό λογιστικό μοντέλο 4PL (Revelle & French, 2013) για διχοτομικά απαντημένα ερωτήματα δίνει την πιθανότητα να απαντηθεί το ερώτημα  $i$  από κάποιον με ικανότητα  $\theta_i$  σε συνάρτηση με τέσσερις παραμέτρους  $\alpha_i$ ,  $\delta_i$ ,  $\gamma_i$ ,  $\zeta_i$  ως εξής:

$$P(x|\theta_i, \alpha_j, \delta_j, \gamma_j, \zeta_j) = \gamma_j + \frac{\zeta_j - \gamma_j}{1 + e^{\alpha_j(\delta_j - \theta_i)}}$$

όπου οι παράμετροι  $\alpha_i$ ,  $\delta_i$ ,  $\gamma_i$  εκφράζουν τη διακριτική ικανότητα (discrimination), τη δυσκολία (difficulty) και την τυχαιότητα (guessing) των απαντήσεων αντίστοιχα και η παράμετρος  $\zeta_i$  είναι μια άνω ασύμπτωτη, που μοντελοποιεί τη τάση να μην απαντηθεί ποτέ σωστά το ερώτημα.

Το διπαραμετρικό λογιστικό μοντέλο (2PL) αποτελεί υποπερίπτωση του τετραπαραμετρικού με  $\gamma_i=0$  και  $\zeta_i=1$ , δηλαδή θεωρεί ότι τα ερωτήματα διαφοροποιούνται μόνο από τη δυσκολία  $\delta_i$  και τη διακριτική τους ικανότητα  $\alpha_i$ .

Στην περίπτωση πολυτομικών απαντήσεων σε μια κλίμακα 1 έως  $n$  δημιουργούνται  $n-1$  κατώφλια, οπότε η πιθανότητα να δοθεί η απάντηση  $k$  είναι συνάρτηση της ικανότητας των κατωφλίων  $k-1$  και  $k$  και των παραμέτρων δυσκολίας  $\delta_i$  και διακριτικής ικανότητας  $\alpha_i$  του ερωτήματος για την περίπτωση 2PL μοντέλου. Το Graded Response Model (GRM) είναι επέκταση του 2PL μοντέλου για πολυτομικά δεδομένα και σε αυτό η πιθανότητα απάντησης της στάθμης  $k$  ενός ερωτήματος δίνεται από τη γενική σχέση (Revelle & French, 2013):

$$\begin{aligned} P(r = k|\theta_i, \delta_k, \delta_{k-1}, \alpha_k) &= \frac{P(r|\theta_i, \delta_{k-1}, \alpha_k)}{1} - \frac{P(r|\theta_i, \delta_k, \alpha_k)}{1} \\ &= \frac{1}{1 + e^{\alpha_k(\delta_{k-1} - \theta_i)}} - \frac{1}{1 + e^{\alpha_k(\delta_k - \theta_i)}} \end{aligned}$$

Στην IRT, το αποτέλεσμα (score) για ένα λανθάνον χαρακτηριστικό δεν εξαρτάται από μια συγκεκριμένη ομάδα ερωτημάτων και οι ιδιότητες των ερωτημάτων μπορούν να μοντελοποιηθούν ανεξάρτητα από τις ιδιότητες των δοκιμασιών και των εξεταζόμενων. Άρα, οι εξεταζόμενοι δεν χρειάζεται να εξεταστούν στην ίδια ομάδα ερωτημάτων, ώστε να λάβουν συγκρίσιμα αποτελέσματα για το ίδιο λανθάνον χαρακτηριστικό. Εάν είναι γνωστές οι παράμετροι των ερωτημάτων, τότε αυτά μπορούν να συνδυαστούν με διαφορετικούς τρόπους, ώστε να παραχθούν διαφορετικές αλλά συγκρίσιμες δοκιμασίες (Shankness, 2014).

Έτσι, η IRT επιτρέπει να σχεδιαστούν δοκιμασίες με συγκεκριμένα ενδογενή χαρακτηριστικά για ένα δεδομένο πληθυσμό εξεταζόμενων (Hambleton & Jones, 1993).

Η συνάρτηση πληροφορίας του ερωτήματος  $j$  (Item Information Function)

υπολογίζεται από τη σχέση:  $I_j(\theta_i) = \frac{[P'_j(\theta_i)]^2}{P_j(\theta_i)(1-P_j(\theta_i))}$ , όπου  $P_j(\theta_i)$  είναι η συνάρτηση

πιθανότητας σωστής απάντησης στην ερώτηση  $j$  και  $P'_j(\theta_i)$  η πρώτη παράγωγός της.

Η συνάρτηση πληροφορίας αντιπροσωπεύει τη συνεισφορά κάθε ερωτήματος στο μετρούμενο λανθάνον χαρακτηριστικό. Ερωτήματα με μεγαλύτερη διακριτική ικανότητα (discrimination) συνεισφέρουν περισσότερο στην ακρίβεια της εξέτασης από ερωτήματα με χαμηλότερη διακριτική ικανότητα και άρα δίνουν υψηλότερη πληροφορία (Ziziba, 2013).

Απλή άθροιση των συναρτήσεων πληροφορίας ερωτήματος για όλα τα ερωτήματα μιας δοκιμασίας θα δώσει τη συνάρτηση πληροφορίας της δοκιμασίας (Test Information Function). Αυτό το χαρακτηριστικό είναι ιδιαίτερα σημαντικό και μοναδικό στα μοντέλα IRT, επειδή επιτρέπει στους ερευνητές να αποφασίζουν εύκολα ποια ερωτήματα θα επιλέξουν, βασιζόμενοι στο αντικείμενο που ενδιαφέρει, αλλά και στην επίδραση των ερωτημάτων στη συνολική πληροφορία της δοκιμασίας. Έτσι, μπορούν να επιλεγούν ερωτήματα που παρουσιάζουν περισσότερη πληροφορία σε συγκεκριμένη περιοχή της κλίμακας της ικανότητας θ ώστε να μεγιστοποιηθεί η ακρίβεια των αποτελεσμάτων στην περιοχή αυτή (Zanon, Hutz, Yoo, & Hambleton, 2016).

#### 2.4. Πρόγραμμα σπουδών και διατύπωση των εκπαιδευτικών στόχων

Ένα γνωστικό αντικείμενο (module) ενός εκπαιδευτικού προγράμματος (curriculum) μπορεί να έχει διαφορετικό εύρος που, κυμαίνεται από μια ολόκληρη γνωστική περιοχή (π.χ. στατιστική), ένα μέρος μιας γνωστικής περιοχής (π.χ. πιθανότητες) έως ένα συγκεκριμένο θέμα που θα μπορούσε να είναι το κεφάλαιο ενός βιβλίου (π.χ. η κανονική κατανομή). Προφανώς, το εύρος κάθε γνωστικού αντικειμένου του εκπαιδευτικού



Εικόνα 1: Δενδροειδής ανάπτυξη των εκπαιδευτικών στόχων ενός γνωστικού αντικειμένου. Η διάσταση  $z$  αντιστοιχεί στο επίπεδο επίτευξης του κάθε στόχου.

προγράμματος καθορίζεται από τον σχεδιαστή του προγράμματος βάσει των αναγκών που πρέπει αυτό να εξυπηρετεί, της φιλοσοφίας του οργανισμού που το παρέχει, των

σκοπών και των στόχων που του αποδίδονται, της δομής και των μεθόδων που επιλέγονται (O'Neill, 2015).

Κάθε γνωστικό αντικείμενο σε ένα εκπαιδευτικό πρόγραμμα θα έχει προσδοκώμενα, αποτελέσματα/στόχους, που αντιστοιχούν σε γνώσεις/δεξιότητες και είναι σε αντιστοιχία με το πρόγραμμα σπουδών. Στην Εικόνα 1, απεικονίζονται οι εκπαιδευτικοί στόχοι να εκτείνονται σε μια δενδροειδή δομή από την ευρύτερη προς την ειδικότερη γνώση στο επίπεδο  $xy$ , ενώ ο βαθμός επίτευξης κάθε στόχου, που αποτελεί το "φύλλο" του παραπάνω δένδρου, το εκτείνει στην τρίτη διάσταση  $z$ .

### 3. Μεθοδολογία

Στην παρούσα εργασία, θεωρούμε την πραγματική ικανότητα ενός εξεταζόμενου σε ένα γνωστικό αντικείμενο (module) ως μια λανθάνουσα μεταβλητή, που μετρίεται σε διαφορετικές χρονικές στιγμές, (α) ενδιαμέσως, δηλαδή αμέσως μετά τη λήξη της διδασκαλίας μιας ενότητας του γνωστικού αντικειμένου και (β) στο τέλος της εκπαιδευτικής περιόδου, όταν ολοκληρώνεται η διδασκαλία του συνόλου του γνωστικού αντικειμένου.

Ως γνωστικό αντικείμενο θεωρείται ένα μέρος από μια γνωστική περιοχή, που διδάσκεται αυτόνομα στο πλαίσιο ενός εκπαιδευτικού ή επιμορφωτικού προγράμματος (π.χ. ένα εξάμηνο πανεπιστημιακό μάθημα, μια θεματική ενότητα του ΕΑΠ, επιμόρφωση για την ανάπτυξη μιας δεξιότητας).

Η μέτρηση της ικανότητας γίνεται με μία σειρά από δοκιμασίες, που αξιολογούν συγκεκριμένους γνωστικούς στόχους, οι οποίοι αντιστοιχούν στις ενότητες που απαρτίζουν το υπό εξέταση γνωστικό αντικείμενο.

Οι δοκιμασίες μπορεί να είναι:

1. Ενδιάμεσα διαγωνίσματα, που αποτελούνται από ερωτήματα κλειστού τύπου με αιτιολόγηση, τα οποία εκτελούνται σε χαρτί και εξετάζουν τους γνωστικούς στόχους μιας ενότητας του γνωστικού αντικειμένου.
2. Ασκήσεις είτε εργαστηριακές, είτε προβλήματα που αντιμετωπίζονται στο σπίτι, πάνω σε συγκεκριμένους στόχους μιας ενότητας του γνωστικού αντικειμένου
3. Τελικά διαγωνίσματα που αξιολογούν εύλογο μέρος των στόχων του υπό αξιολόγηση γνωστικού αντικειμένου, ανάλογα με τους διαθέσιμους πόρους.

Το προτεινόμενο πλαίσιο διαδικασίας περιλαμβάνει χρονολογικά τα παρακάτω βήματα:

1. Προσδιορισμό της δομής του περιεχομένου του γνωστικού αντικειμένου.
2. Διατύπωση των εκπαιδευτικών στόχων ανά ενότητα του γνωστικού αντικειμένου.
3. Ενδιάμεσες δοκιμασίες σε ομάδες εκπαιδευτικών στόχων/ενοτήτων, που παρουσιάζουν λογική συνάφεια και εξαρτήσεις μεταξύ τους και μπορούν να αντιμετωπιστούν ως ευρύτερες ενότητες. Στο σύνολο των ενδιάμεσων δοκιμασιών εξετάζονται όλοι οι εκπαιδευτικοί στόχοι, που διατυπώθηκαν στο προηγούμενο βήμα.
4. Ανάλυση των αποτελεσμάτων των ενδιάμεσων δοκιμασιών του βήματος 3 για την επιλογή των δοκιμασιών της τελικής εξέτασης.

#### 3.1. Προσδιορισμός της δομής περιεχομένου του γνωστικού αντικειμένου

Τα περιεχόμενα ενός γνωστικού αντικειμένου, όπως ορίστηκε παραπάνω για τους σκοπούς αυτής της έρευνας, περιλαμβάνουν ένα μέρος από μια γνωστική περιοχή και χωρίζονται συνήθως σε ενότητες, που αντιστοιχούν στα κεφάλαια ενός παραδοσιακού βιβλίου. Κάθε μια από τις ενότητες περιγράφει συγκεκριμένες έννοιες

ή μεθόδους από ένα πολύ συγκεκριμένο μέρος του γνωστικού αντικειμένου, το οποίο κρίνεται ως απαραίτητη γνώση για τους εξεταζόμενους.

Οι ενότητες έχουν συνήθως σχέσεις αλληλεξάρτησης, που δηλώνουν προ-απαίτηση τόσο μεταξύ τους, όσο και με ενότητες άλλων μαθημάτων, που προηγούνται ή ακολουθούν.

Η απεικόνιση των αλληλεξαρτήσεων των ενοτήτων μπορεί γίνει να με τη μορφή ενός κατευθυνόμενου γράφου\*. Από την απεικόνιση της δομής του γνωστικού αντικειμένου μπορεί να γίνουν φανερές τυχόν ομαδοποιήσεις ενοτήτων (στο παράδειγμα αναπαρίστανται με ίδια χρώματα), που προκύπτουν από τη λογική σύνδεση των ενοτήτων και οι οποίες μπορούν να αποτελέσουν αντικείμενα κοινής ενδιάμεσης αξιολόγησης με την ίδια ενδιάμεση δοκιμασία.

### **3.2. Ενδιάμεση αξιολόγηση (Interim assessment)**

Η ενδιάμεση αξιολόγηση των εξεταζόμενων προβλέπει μια ομάδα ενδιάμεσων δοκιμασιών, που εκτελούνται κατά τη διάρκεια του εκπαιδευτικού προγράμματος, παράλληλα με τη διδασκαλία.

Φαίνεται αυτονόητο ότι όσο μεγαλύτερη είναι η διαμέριση του γνωστικού αντικειμένου, τόσο λεπτομερέστερη θα είναι η αποτύπωση της εικόνας των επιτευγμάτων των εξεταζόμενων. Για πρακτικούς, όμως, λόγους (χρόνου και διαθέσιμων πόρων) η ενδιάμεση αξιολόγηση γίνεται σε εκπαιδευτικούς στόχους ομάδων ενοτήτων, που παρουσιάζουν λογικές εξαρτήσεις ως προς το περιεχόμενό τους.

Στο προτεινόμενο πλαίσιο διαδικασίας οι ενδιάμεσες δοκιμασίες μπορούν να είναι ενδιάμεσα τεστ, ασκήσεις ή άλλες αξιολογούμενες δραστηριότητες αρκεί να μπορούν να βαθμολογηθούν διχοτομικά ως επιτυχείς/ανεπιτυχείς. Στην παρούσα έρευνα χρησιμοποιούνται ως ενδιάμεσες δοκιμασίες, ενδιάμεσα τεστ που περιλαμβάνουν μια ομάδα ερωτημάτων (π.χ. 12-14) πολλαπλής επιλογής, τα οποία πρέπει να απαντηθούν σε συγκεκριμένο χρόνο (π.χ. μια ώρα). Οι εξεταζόμενοι επιλέγουν τη σωστή απάντηση και κατά προτίμηση την αιτιολογούν προκειμένου να διαπιστωθεί, ή όχι, η γνώση του συγκεκριμένου στοιχείου, που εξετάζεται και να αποθαρρυνθούν οι τυχαίες απαντήσεις. Οι ενδιάμεσες δοκιμασίες είναι όσες και οι ομάδες ενοτήτων, που προκύπτουν από τον προσδιορισμό της δομής του γνωστικού αντικειμένου και διεξάγονται με τον παραδοσιακό τρόπο σε χαρτί (Paper Based Assessment) και ενιαία για όσους επιθυμούν να συμμετάσχουν. Οι περιεχόμενες ερωτήσεις επιλέγονται αντιπροσωπευτικά, ώστε να αξιολογούνται οι εκπαιδευτικοί στόχοι, που αντιστοιχούν σε κάθε ομάδα ενοτήτων.

Η φύση των ερωτημάτων πολλαπλής επιλογής, που περιλαμβάνονται στο διαγώνισμα, είναι τέτοια, ώστε να απαιτεί κατά το δυνατόν ανώτερες νοητικές λειτουργίες και οι εσφαλμένες επιλογές ("δολώματα"), οι οποίες προσφέρονται να αντιστοιχούν σε συνήθη λάθη των εκπαιδευόμενων (Lin & Singh, 2013; Scott, Stelzer, & Gladding, 2006).

Αντίστοιχα, θα μπορούσαν να δίνονται στους εκπαιδευόμενους ομάδες από ασκήσεις, που θα εκτελούνταν είτε στο εργαστήριο, είτε ως ενδιάμεσες δραστηριότητες στο σπίτι και θα εξέταζαν περισσότερο εργαστηριακές δεξιότητες-στόχους της αντίστοιχης ενότητας. Κάθε άσκηση αξιολογείται είτε διχοτομικά ως προς την επιτυχία ή την αποτυχία, είτε πολυτομικά ως προς την επίτευξη των συνδεδεμένων με αυτή στόχων της αντίστοιχης ενότητας.

### 3.3. Αξιολόγηση των ενδιάμεσων δοκιμασιών - επιλογή ερωτήσεων τελικής δοκιμασίας

#### 3.3.1. Γενικά χαρακτηριστικά τελικής εξέτασης

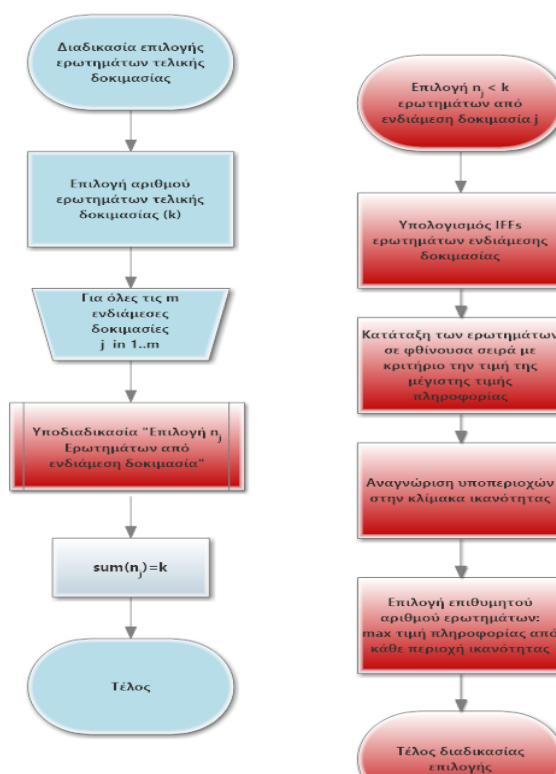
Η τελική αξιολόγηση για το γνωστικό αντικείμενο γίνεται με μια τελική δοκιμασία, που αποτελείται από ομάδες ερωτημάτων, οι οποίες διαχωρίζουν την κλίμακα ικανότητας σε τμήματα. Προφανής διαχωρισμός είναι σε τρία τμήματα που αντιστοιχούν σε: εύκολα, μέτριας δυσκολίας και δύσκολα ερωτήματα. Όλα τα ερωτήματα, που αποτελούν την τελική δοκιμασία, προέρχονται από αναδιατύπωση ερωτημάτων, τα οποία έχουν ήδη χρησιμοποιηθεί στις ενδιάμεσες δοκιμασίες, έχοντας μετατραπεί σε ψευδο-ανοικτού τύπου, δηλαδή δίνονται απαντημένα με τυχαίο συνδυασμό, σωστά ή λάθος, και ο εξεταζόμενος καλείται να βαθμολογήσει την κάθε απάντηση και να αιτιολογήσει τη βαθμολογία του. Στόχος είναι οι εξεταζόμενοι να αναλύσουν πλήρως το πρόβλημα κάθε ερωτήματος και να εξηγήσουν επαρκώς τα συστατικά της λύσης του (για παράδειγμα μπορούν να αποδείξουν και ότι κάποιες από τις προτεινόμενες απαντήσεις είναι άτοπες επειδή προσκρούουν σε δεδομένες θεωρητικές προβλέψεις), προκειμένου να εκτιμηθεί ο βαθμός αναλυτικής σκέψης και κριτικής ικανότητας.

Η βαθμολογία κάθε ερωτήματος γίνεται πολυτομικά με κατάταξη, στα επίπεδα: 1- Απομνημόνευση, 2-Κατανόηση, 3-Εφαρμογή, 4-Ανάλυση και 5-Κρίση/Αξιολόγηση. Σημαντική παράμετρος είναι η συνεπής εφαρμογή των προαποφασισμένων κριτηρίων διόρθωσης από τον διδάσκοντα, ενώ, στην περίπτωση περισσότερων του ενός αξιολογητών, θα πρέπει να υπάρχει τόσο εκ των προτέρων συμφωνία ως προς τα κριτήρια αξιολόγησης κάθε ερωτήματος, όσο και εφαρμογή τεχνικών, που θα επιτρέψουν συναντίληψη μεταξύ τους (π.χ. δοκιμαστική κοινή βαθμολογία αντιπροσωπευτικού αριθμού δοκιμασιών).

#### 3.3.2. Αλγόριθμος Επιλογής ερωτημάτων

Ο αλγόριθμος για την επιλογή των ερωτημάτων της τελικής δοκιμασίας, που προτείνεται, περιλαμβάνει τα παρακάτω βήματα και συνοψίζεται στα διαγράμματα ροής στην Εικόνα 2:

1. Επιλογή του αριθμού  $k$  των ερωτημάτων της τελικής εξέτασης. Αυτός θα εξαρτάται από τους διαθέσιμους πόρους και το διαθέσιμο χρόνο
2. Για κάθε μια από τις  $m$  ενδιάμεσες δοκιμασίες  $i$  επιλέγεται  $n_{ij}$  αριθμός ερωτημάτων, ώστε  $\sum_{i=1}^m n_{ij} = k$  ως εξής:
  - a. Κατατάσσονται τα ερωτήματα κάθε ενδιάμεσης δοκιμασίας  $i$  σε φθίνουσα σειρά ανάλογα με την μέγιστη τιμή της πληροφορίας (ή ισοδύναμα την διακριτική ικανότητά τους).
  - b. Αναγνωρίζονται οι υποπεριοχές στο εύρος



Εικόνα 2: Διάγραμμα ροής αλγορίθμου επιλογής ερωτημάτων τελικής δοκιμασίας

ικανότητας  $\theta$ , στις οποίες εμπίπτουν τα μέγιστα των συναρτήσεων πληροφορίας των ερωτημάτων. Ο αριθμός των υποπεριοχών είναι επιλέξιμη παράμετρος (π.χ. τρία: Εύκολα, Μέτρια, Δύσκολα)

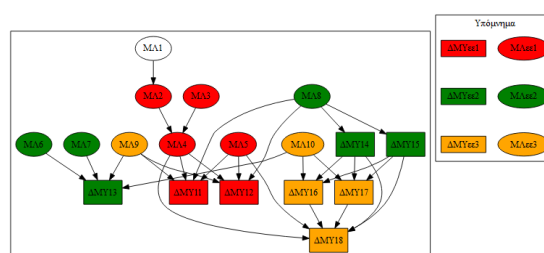
- c. Επιλέγεται για την τελική δοκιμασία ίσος περίπου αριθμός  $n_{ij}$  ερωτημάτων από κάθε ενδιάμεση δοκιμασία, που ταυτόχρονα ικανοποιούν τις συνθήκες:
  - i. έχουν τη μεγαλύτερη δυνατή τιμή μέγιστης πληροφορίας
  - ii. ισοκατανέμονται στις ενδιάμεσες δοκιμασίες
  - iii. ισοκατανέμονται στις υποπεριοχές εύρους ικανότητας

## 4. Εφαρμογή

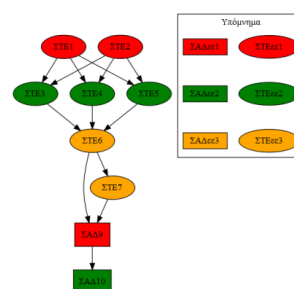
### 4.1. Προσδιορισμός της δομής περιεχομένου του γνωστικού αντικειμένου και διατύπωση των εκπαιδευτικών στόχων ανά ενότητα

Το προτεινόμενο πλαίσιο διαδικασίας εφαρμόστηκε σε δυο πανεπιστημιακά μαθήματα του Τμήματος Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πατρών το ακαδημαϊκό έτος 2016-2017 ως εξής:

- Στο μάθημα Στατιστική Ανάλυση Δεδομένων (ΣΑΔ), μάθημα του Γ' εξαμήνου, που έχει ως προαπαιτούμενο το μάθημα Στατιστική των επιχειρήσεων (ΣΤΕ): Στην τελική δοκιμασία μελετάται πληθυσμός 274 εξεταζόμενων, εκ των οποίων 73 σε ενδιάμεσες δοκιμασίες αξιολόγησης.
- Στο μάθημα Μαθηματικός Λογισμός (ΜΛ), μάθημα Α' εξαμήνου σπουδών, που αποτελεί προαπαιτούμενο του μαθήματος Δυναμικά Μαθηματικά Υποδείγματα (ΔΜΥ) του Β' εξαμήνου σπουδών: Στην τελική δοκιμασία μελετάται πληθυσμός 394 φοιτητών, εκ των οποίων 113 συμμετείχαν σε ενδιάμεσες δοκιμασίες.



Εικόνα 3: Εξαρτήσεις ενότητων σε ΜΛ και ΔΜΥ



Εικόνα 4: Εξαρτήσεις ενότητων σε μαθήματα ΣΑΔ και ΣΤΕ

Η ανάλυση εξαρτήσεων των ενότητων των μαθημάτων αναπαρίσταται στους κατευθυνόμενους γράφους των Εικόνων 3 και 4. Στις Εικόνες 3 και 4 αναλύεται επιπλέον η δομή ενότητων και των μαθημάτων Στατιστική των Επιχειρήσεων (ΣΤΕ) και Δυναμικά Μαθηματικά Υποδείγματα (ΔΜΥ), τα οποία όμως δεν αποτελούν πεδίο εφαρμογής του περιγραφόμενου πλαισίου μεθοδολογίας στην παρούσα εργασία.

Με ίδια χρώματα αναπαρίστανται οι ενότητες που εξετάζονται με ίδια ενδιάμεση δοκιμασία: κόκκινο - 1η ενδιάμεση δοκιμασία, πράσινο - 2η ενδιάμεση δοκιμασία και πορτοκαλί - 3η ενδιάμεση δοκιμασία.

Οι εκπαιδευτικοί στόχοι διατυπώνονται και καταγράφονται ανά ενότητα, προκειμένου να αξιολογηθούν σε επίπεδα, σύμφωνα με την αναμορφωμένη ταξινόμια Bloom, η οποία ταιριάζει καλύτερα στο πεδίο εφαρμογής (πανεπιστημιακό μάθημα), όπου η αξιολόγηση γίνεται μόνο στον γνωστικό τομέα.

**Πίνακας 1:** Ενδεικτικοί στόχοι ενότητας μαθήματος

Μάθημα: Στατιστική Ανάλυση Δεδομένων
Ενότητα 10: Απλή Γραμμική Παλινδρόμηση και Συσχέτιση
Στο τέλος του κεφαλαίου θα πρέπει να:
αναγνωρίζετε πότε η παλινδρόμηση είναι χρήσιμη για την περίπτωση μας
κατασκευάζετε ένα μοντέλο παλινδρόμησης
υπολογίζετε την εξίσωση παλινδρόμησης
υπολογίζετε τη συνδιακύμανση και τη συσχέτιση δύο τυχαίων μεταβλητών
υπολογίζετε διαστήματα εμπιστοσύνης για τους συντελεστές παλινδρόμησης
εκτελείτε έλεγχο υπόθεσης για τους συντελεστές της παλινδρόμησης
εφαρμόζετε έλεγχο ANOVA με τα αποτελέσματα της παλινδρόμησης
αναλύετε τα κατάλοιπα ώστε να ελέγξετε υποθέσεις αναφορικά με την ορθότητα της παλινδρόμησης
εκτελείτε παλινδρόμηση με το R
εφαρμόζετε την έννοια της συνδιακύμανσης για τη γραμμική σύνθεση τυχαίων μεταβλητών

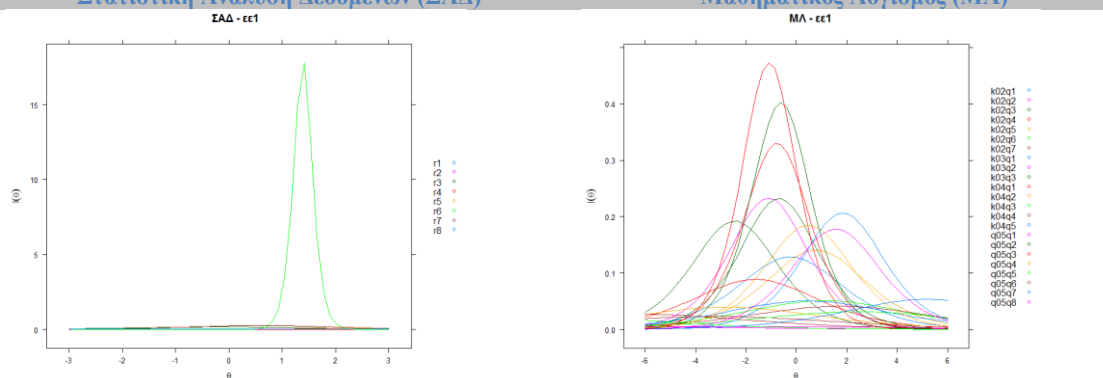
Ενδεικτικά αναφέρονται στον Πίνακα 1, οι εκπαιδευτικοί στόχοι, για την ενότητα 10 του μαθήματος ΣΑΔ.

#### 4.2. Επιλογή Ερωτημάτων τελικής εξέτασης από τα ερωτήματα των ενδιάμεσων δοκιμασιών

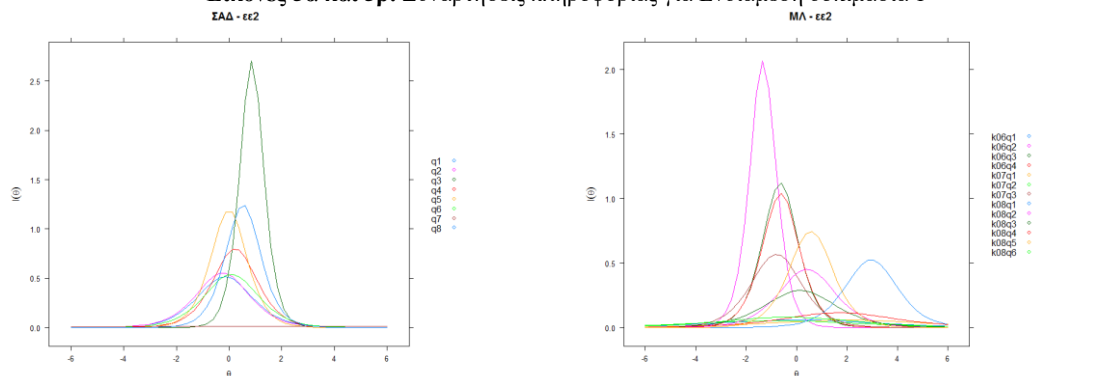
Στην παρούσα έρευνα, η ανάλυση των δοκιμασιών, προκειμένου να προσδιοριστούν οι παράμετροι των ερωτημάτων των ενδιάμεσων, της τελικής δοκιμασίας και οι συναρτήσεις πληροφορίας τους, γίνεται με 2PL IRT μοντέλο για τα διχοτομικά δεδομένα και με GRM μοντέλο για τα πολυτομικά δεδομένα. Ως εργαλεία χρησιμοποιήθηκαν η γλώσσα στατιστικής επεξεργασίας R και τελικά το πακέτο IRT ανάλυσης mirt (Chalmers, 2012), ενώ δοκιμάστηκαν, επίσης, τα αντίστοιχα IRT πακέτα ltm (Rizouros, 2006), psych (Revelle, 2014) με παρόμοια αποτελέσματα.

Στις εικόνες 5 και 6 φαίνονται, ενδεικτικά, οι συναρτήσεις πληροφορίας  $I(\theta)$  για την πρώτη (Εικόνα 5α) και τη δεύτερη (Εικόνα 5β) ενδιάμεση δοκιμασία στα μαθήματα Στατιστική Ανάλυση Δεδομένων και Μαθηματικός Λογισμός

**Στατιστική Ανάλυση Δεδομένων (ΣΑΔ)** **Μαθηματικός Λογισμός (ΜΑ)**



**Εικόνες 5α και 5β:** Συναρτήσεις πληροφορίας για Ενδιάμεση δοκιμασία 1



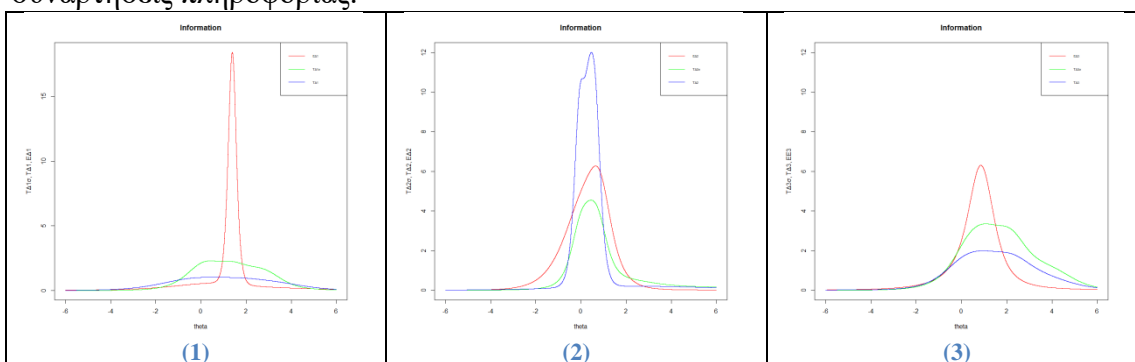
**Εικόνες 6α και 6β:** Συναρτήσεις πληροφορίας για Ενδιάμεση δοκιμασία 2

Η μέγιστη τιμή της συνάρτησης πληροφορίας για κάθε ερώτημα των ενδιάμεσων δοκιμασιών και η τιμή του  $\theta$ , στην οποία αυτή παρατηρείται κατά προσέγγιση, μπορεί να υπολογιστεί, ώστε να επιλεγούν τα ερωτήματα της τελικής δοκιμασίας. Στον Πίνακα 2, παρουσιάζεται, ενδεικτικά, αυτός ο υπολογισμός για τις ενδιάμεσες δοκιμασίες 1 και 2 των δυο μαθημάτων.

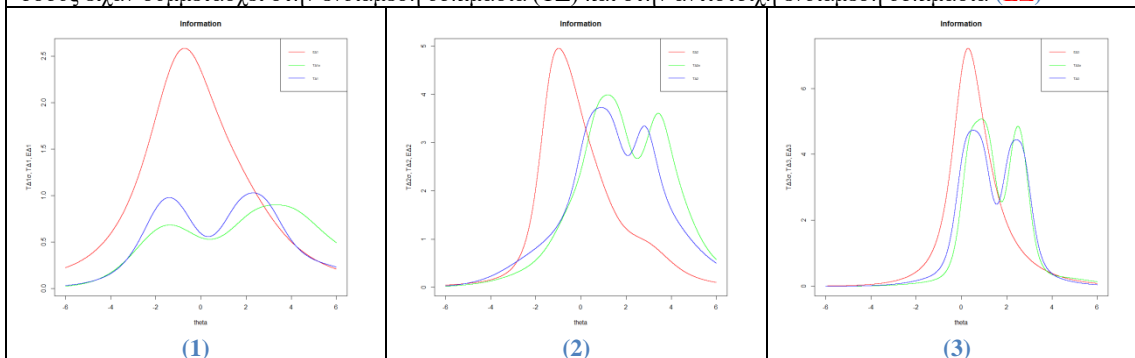
	Στατιστική Ανάλυση Δεδομένων (ΣΑΔ)	Μαθηματικός Λογισμός (ΜΛ)	Πίνακας 2
Ενδιάμεση Δοκιμασία 1	<pre>&gt; data.info.pr1       min I(θ)      max I(θ)      θ [1,] 4.134050e-04 4.368281e-04 -4.0 [2,] 1.785105e-03 5.365156e-03 -4.0 [3,] 2.182918e-02 1.823579e-01  0.0 [4,] 5.135840e-03 1.529839e-02  4.0 [5,] 3.490092e-03 9.848925e-02  3.5 [6,] 4.520843e-14 1.787393e+01  1.4 [7,] 8.075904e-03 2.836167e-01  0.6 [8,] 2.059748e-02 5.040003e-02  0.5</pre>	<pre>&gt; data.info.pr1       min I(θ)      max I(θ)      θ [1,] 0.0057567793 0.128553543 -0.233 [2,] 0.0009732607 0.233269641 -1.105 [3,] 0.0004679861 0.192423334 -2.439 [4,] 0.0036882793 0.088964150 -1.625 [5,] ..... [16,] 0.0007104220 0.005821734 -6.000 [17,] 0.0003643725 0.402583850 -0.614 [18,] 0.0001120336 0.472833838 -1.078 [19,] 0.0028087162 0.185025376  0.470 [20,] 0.0004843917 0.016704344 -6.000 [23,] 0.0011989240 0.178134064  1.563</pre>	
Ενδιάμεση Δοκιμασία 2	<pre>&gt; data.info.pr3       min I(θ)      max I(θ)      θ [1,] 2.593926e-02 0.1234780565  0.0 [2,] 9.054084e-03 0.2754831019  0.6 [3,] 2.431614e-03 0.4832382123  0.8 [4,] 1.998770e-06 2.5705710263  0.8 [5,] 1.347367e-02 0.0632335971  1.6 [6,] 8.374470e-04 0.8128507372  0.6 [7,] 8.789168e-04 0.0009321612  4.0 [8,] 4.729659e-06 2.0976203430  1.0</pre>	<pre>&gt; data.info.pr3       min I(θ)      max I(θ)      θ [1,] 7.682209e-04 0.3112371 -0.623 [2,] 8.423735e-05 0.4756000  1.268 [3,] 2.400045e-05 0.7839894  0.652 [4,] 9.504777e-06 0.9463317  0.628 [5,] 1.095144e-05 0.9108287  0.661 [6,] 1.153489e-06 1.6140743  0.115 [7,] 5.988700e-08 2.3081828  0.205 [11,] ..... [11,] 4.283891e-03 0.1227562  0.741</pre>	

### 4.3. Σύγκριση τελικών και ενδιάμεσων δοκιμασιών

Η ανάλυση της τελικής δοκιμασίας με το μοντέλο GRM (δεδομένου ότι πρόκειται για πολυτομικά δεδομένα) και των ενδιάμεσων δοκιμασιών στις ενότητες με το μοντέλο δύο παραμέτρων (2PL) δίνει την συνάρτηση πληροφορίας, ανά εξεταζόμενη ενότητα, για την τελική δοκιμασία όλων των εξεταζόμενων (ΤΔσ), την τελική δοκιμασία των εξεταζόμενων, που συμμετείχαν στις ενδιάμεσες δοκιμασίες (ΤΔ) και τις ενδιάμεσες δοκιμασίες (ΕΔ). Στις Εικόνες 7 και 8 απεικονίζονται οι παραπάνω συναρτήσεις πληροφορίας.



Εικόνα 7: ΣΑΔ - Συναρτήσεις πληροφορίας στην τελική δοκιμασία για το σύνολο των εξεταζόμενων (ΤΔσ), για όσους είχαν συμμετάσχει στην ενδιάμεση δοκιμασία (ΤΔ) και στην αντίστοιχη ενδιάμεση δοκιμασία (ΕΔ)



Εικόνα 8: ΜΛ - Συναρτήσεις πληροφορίας στην τελική δοκιμασία για το σύνολο των εξεταζόμενων (ΤΔσ), για όσους είχαν συμμετάσχει στην ενδιάμεση δοκιμασία (ΤΔ) και στην αντίστοιχη ενδιάμεση δοκιμασία (ΕΔ)

Το εμβαδόν  $S$  κάτω από την κάθε συνάρτηση πληροφορίας είναι μέτρο της ολικής πληροφορίας στο διάστημα ικανότητας που παρέχει η κάθε δοκιμασία και υπολογίζεται στον Πίνακα 3

**Πίνακας 3:** Εμβαδό κάτω από τη συνάρτηση πληροφορίας και αριθμός ερωτημάτων ανά δοκιμασία

Δοκιμασία	ΣΑΔ			ΜΛ		
	Εικόνα	S	N	Εικόνα	S	N
ΕΔ1	7.1	11,73	8	8.1	14,48	23
ΤΔ1σ	7.1	9,67	3	8.1	7,61	3
ΤΔ1	7.1	6,28	3	8.1	7,12	3
ΕΔ2	7.2	14,08	8	8.2	16,77	13
ΤΔ2σ	7.2	9,56	3	8.2	20,56	5
ΤΔ2	7.2	15,39	3	8.2	19,07	5
ΕΔ3	7.3	11,57	8	8.3	17,07	11
ΤΔ3σ	7.3	13,69	4	8.3	15,03	2
ΤΔ3	7.3	9,27	4	8.3	15,46	2

Όπου: Με ΤΔσ αναφέρεται η τελική δοκιμασία για το συνολικό εξεταζόμενο πληθυσμό, ενώ με ΤΔ μόνο για όσους συμμετείχαν και στην ενδιάμεση αξιολόγηση -  $S$  το εμβαδό κάτω από τη συνάρτηση πληροφορίας της δοκιμασίας -  $N$  ο αριθμός των ερωτημάτων της δοκιμασίας

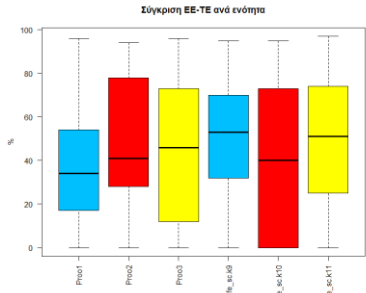
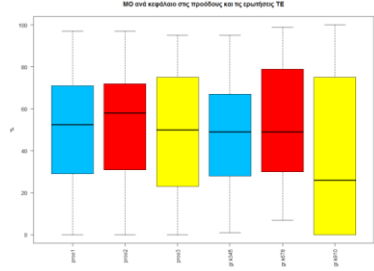
Όπως φαίνεται στις Εικόνες 7.3, 8.2, 7.2 & 8.3 και στον Πίνακα 3, σε τέσσερις από τις έξι ενότητες, των υπό μελέτη μαθημάτων, η πληροφορία των τελικών δοκιμασιών ανά ενότητα είναι είτε μεγαλύτερη, είτε αρκετά κοντά στην πληροφορία που δίνουν οι ενδιάμεσες δοκιμασίες, παρόλο που ο αριθμός των ερωτημάτων, τα οποία περιέχουν οι τελικές δοκιμασίες είναι κατά πολύ μικρότερος (μικρότερος από το μισό) από τον αριθμό των ερωτημάτων των αντίστοιχων ενδιάμεσων δοκιμασιών. Οι δυο περιπτώσεις, που η πληροφορία της ενδιάμεσης δοκιμασίας είναι διπλάσια της τελικής (Εικόνες 7.1 & 8.1), μπορούν να αποδοθούν στους κατά πολύ μεγαλύτερους αριθμούς ερωτημάτων αυτών των ενδιάμεσων δοκιμασιών έναντι του αριθμού των ερωτημάτων στην τελική δοκιμασία. Σε όλες τις περιπτώσεις, αν ληφθεί υπόψη ο αριθμός των ερωτημάτων, που δίνουν την αντίστοιχη συνάρτηση πληροφορίας, η τελική δοκιμασία παρουσιάζει αύξηση της πληροφορίας σε σχέση με την ενδιάμεση. Επιπλέον, σε όλες τις περιπτώσεις, οι συναρτήσεις πληροφορίας των τελικών δοκιμασιών παρουσιάζουν υψηλότερες τιμές, που κατανομούνται σε μεγαλύτερο εύρος της κλίμακας ικανότητας, σε αντίθεση με τις αντίστοιχες συναρτήσεις πληροφορίας των ενδιάμεσων δοκιμασιών, που παρουσιάζουν οξεία μέγιστα σε συγκεκριμένες τιμές της ικανότητας, δίνοντας μεγαλύτερη ακρίβεια μέτρησης σε μεγαλύτερο εύρος ικανότητας.

Τέλος, η σύγκριση των συναρτήσεων πληροφορίας για την τελική δοκιμασία μεταξύ των εξεταζόμενων, που συμμετείχαν στις ενδιάμεσες δοκιμασίες (ΤΔσ) και όλων των εξεταζόμενων (ΤΔσ), δείχνει ότι μόνο σε 2 περιπτώσεις (Εικόνες 7.2 και 8.1) το εμβαδό κάτω από τις συναρτήσεις πληροφορίας, όσων συμμετείχαν στις ενδιάμεσες δοκιμασίες, είναι μεγαλύτερο από το αντίστοιχο για το σύνολο των εξεταζόμενων.

#### 4.4. Εξαγωγή βαθμολογίας στις ενότητες σε ενδιάμεση και τελική δοκιμασία

Η βαθμολογία επί τοις εκατό για κάθε εξεταζόμενο είναι η ποσοστιαία θέση του στην τυποποιημένη κανονική κατανομή. Η εξαγωγή της βαθμολογίας μπορεί να γίνει συνολικά για μια δοκιμασία ή τμηματικά για επιμέρους ομάδες ερωτημάτων.

Στις Εικόνες 9 και 10 συγκρίνονται τα σκορ των εξεταζόμενων στις ενδιάμεσες δοκιμασίες και στην αντίστοιχη ενότητα της τελικής δοκιμασίας.

Στατιστική Ανάλυση δεδομένων (ΣΑΔ)	Μαθηματικός Λογισμός (ΜΛ)
<p style="text-align: center;"><b>Σύγκριση ΕΕ-ΤΕ ανά ενότητα</b></p>  <p><b>Εικόνα 9:</b> Σύγκριση επιμέρους βαθμολογιών της τελικής δοκιμασίας ανά ενότητα με τις βαθμολογίες στις αντίστοιχες ενδιάμεσες δοκιμασίες</p>	<p style="text-align: center;"><b>ΜΕ ανά κλάδο στις προόδους και τις τριμήνες ΤΕ</b></p>  <p><b>Εικόνα 10:</b> Σύγκριση επιμέρους βαθμολογιών της τελικής δοκιμασίας ανά ενότητα με τις βαθμολογίες στις αντίστοιχες ενδιάμεσες δοκιμασίες</p>
<pre>&gt; t.test(my.data\$fe_sc.k9, my.data\$Proo1, paired=T) Paired t-test data: my.data\$fe_sc.k9 and my.data\$Proo1 t = 3.013, df = 73, p-value = 0.003554 alternative hypothesis: true difference in means is not equal to 0 mean of the differences 12.59459 &gt; t.test(my.data\$fe_sc.k10, my.data\$Proo2, paired=T) Paired t-test data: my.data\$fe_sc.k10 and my.data\$Proo2 t = -1.9077, df = 73, p-value = 0.06036 alternative hypothesis: true difference in means is not equal to 0 &gt; t.test(my.data\$fe_sc.k11, my.data\$Proo3, paired=T) Paired t-test data: my.data\$fe_sc.k11 and my.data\$Proo3 t = 0.70017, df = 73, p-value = 0.486 alternative hypothesis: true difference in means is not equal to 0</pre>	<pre>&gt; t.test(my.data\$proo1, my.data\$gr.k345, paired = TRUE) Paired t-test data: my.data\$proo1 and my.data\$gr.k345 t = 0.66685, df = 113, p-value = 0.5062 alternative hypothesis: true difference in means is not equal to 0 &gt; t.test(my.data\$proo2, my.data\$gr.k678, paired = TRUE) Paired t-test data: my.data\$proo2 and my.data\$gr.k678 t = -0.20294, df = 113, p-value = 0.8395 alternative hypothesis: true difference in means is not equal to 0 &gt; t.test(my.data\$proo3, my.data\$gr.k910, paired = TRUE) Paired t-test data: my.data\$proo3 and my.data\$gr.k910 t = 1.8138, df = 113, p-value = 0.07237 alternative hypothesis: true difference in means is not equal to 0</pre>

Η σύγκριση των βαθμολογιών, που δίνει η μέθοδος στις ενδιάμεσες δοκιμασίες και στις αντίστοιχες ενότητες των διαγωνισμάτων τελικής, εξέτασης δίνει μη στατιστικά σημαντικές διαφορές στους μέσους όρους στις πέντε από τις έξι περιπτώσεις, που εξετάστηκαν. Σε μία περίπτωση (ενότητα 1 του μαθήματος ΣΑΔ, ΣΑΔ9 στην Εικόνα 3) η μέση τιμή των βαθμολογιών στην ενδιάμεση εξέταση είναι χαμηλότερη από την αντίστοιχη μέση τιμή στην τελική εξέταση και η διαφορά είναι στατιστικά σημαντική ( $p < 0.05$ ). Αυτό θα μπορούσε να ερμηνευτεί από το ότι η αντίστοιχη ενότητα είναι προαπαιτούμενη γνώση (όπως φαίνεται στον αντίστοιχο γράφο της Εικόνας 3) για τις άλλες δυο εξεταζόμενες ενότητες και συνεπώς οι εξεταζόμενοι αύξησαν κατά μέσο όρο την ικανότητά τους σε αυτή, προκειμένου να αντιμετωπίσουν τις επόμενες ενότητες.

## 5. Συμπεράσματα - Συζήτηση

Το μέγιστο της συνάρτησης πληροφορίας ερωτήματος και η θέση του στην κλίμακα της ικανότητας  $\theta$ , όπου αυτό παρουσιάζεται είναι αντικειμενικά μέτρα της πληροφορίας που παρέχει ένα ερώτημα. Η μορφή της συνάρτησης πληροφορίας για όλα τα ερωτήματα μιας δοκιμασίας και το εμβαδό κάτω από αυτή αξιολογούν την ακρίβεια της μέτρησης της ικανότητας των εξεταζόμενων με τη συγκεκριμένη δοκιμασία. Στην παρούσα εργασία, χρησιμοποιήσαμε αυτά τα χαρακτηριστικά της συνάρτησης πληροφορίας για να διατυπώσουμε ένα πλαίσιο διαδικασίας επιλογής ερωτημάτων για την τελική αξιολόγηση, από τη δεξαμενή ερωτημάτων που χρησιμοποιήθηκαν στις ενδιάμεσες δοκιμασίες. Δείξαμε ότι η πληροφορία της τελικής δοκιμασίας για τους ίδιους εκπαιδευτικούς στόχους, αυξάνεται αν στην τελική δοκιμασία συμπεριληφθούν αναδιατυπωμένα ερωτήματα των ενδιάμεσων δοκιμασιών, τα οποία έχουν επιλεγεί με κύριο κριτήριο την μέγιστη τιμή πληροφορίας και δευτερεύοντα κριτήρια την κάλυψη ολόκληρης της περιοχής της

κλίμακας ικανότητας θ και την κάλυψη κατά το δυνατόν περισσότερων εκπαιδευτικών στόχων.

Η αύξηση της πληροφορίας στην τελική δοκιμασία μπορεί να αποδοθεί αφενός στον προτεινόμενο τρόπο επιλογής των ερωτημάτων, αφετέρου στην πολυτομική βαθμολόγηση των απαντήσεων των εξεταζόμενων μέσω κλίμακας εκτίμησης του επιπέδου επίτευξης των αντίστοιχων εκπαιδευτικών στόχων. Το ποσοστό συνεισφοράς των δυο παραπάνω παραγόντων στην αύξηση της πληροφορίας της τελικής δοκιμασίας είναι θέμα προς περαιτέρω διερεύνηση.

Οι κατά μέσο όρο βαθμολογίες, που δίνει η μέθοδος, για τους εξεταζόμενους στις ενότητες της τελικής δοκιμασίας και στις ενδιάμεσες δοκιμασίες δεν διαφέρουν σημαντικά στις περισσότερες περιπτώσεις, ώστε να έχει βάση ο ισχυρισμός ότι η εξέταση με τα επιλεγμένα ερωτήματα, σύμφωνα με το προτεινόμενο πλαίσιο διαδικασίας, αξιολογεί με ισοδύναμο τρόπο την επίτευξη των γνωστικών στόχων της κάθε ενότητας. Επιπλέον, οι εξαγόμενες βαθμολογίες με τη χρήση της θεωρίας IRT και του προτεινόμενου πλαισίου διαδικασίας δίνουν συγκρισιμότητα μεταξύ εξεταζόμενων διαφορετικών κύκλων του ίδιου γνωστικού αντικείμενου, δεδομένου ότι η τελική εξέταση διεξάγεται πάντα με δοκιμασίες, που έχουν μεγάλο βαθμό πληροφορίας.

Επεκτάσεις του προτεινόμενου πλαισίου θα μπορούσαν να προκύψουν: α) από τη διατύπωση εκπαιδευτικών στόχων, σύμφωνα με τη ταξινόμια Fink (Fink, 2003), όπου αυτοί εκτείνονται πέρα από τη γνωστική και στην ανθρώπινη διάσταση και συνεπώς θα πρέπει να διαφοροποιηθούν κατάλληλα τα διαγνωστικά μέσα, ώστε να μπορούν να διαγνώσουν αλλαγές σε λανθάνουσες μεταβλητές, όπως για παράδειγμα στα συναισθήματα, στις αξίες, στα ενδιαφέροντα και στην κριτική ικανότητα των εκπαιδευόμενων και β) αν η ικανότητα των εξεταζόμενων σε ένα αντικείμενο δεν θεωρηθεί ως μονοδιάστατη, αλλά ότι εξαρτάται από περισσότερους παράγοντες, όπως η προηγούμενη γνώση/εμπειρία και η προσπάθεια που καταβάλλεται από τον εξεταζόμενο κατά τη διάρκεια της εκπαιδευτικής περιόδου (στην περίπτωση αυτή θα πρέπει να χρησιμοποιηθούν μοντέλα πολυπαραγοντικής IRT).

## 6. Βιβλιογραφία

- Align Assessments with Objectives. (2017, June 3). Retrieved from <http://www.cmu.edu/teaching/assessment/howto/basics/objectives.html>
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing*, Abridged Edition.
- Bang-Jensen, J., & Gutin, G. Z. (2008). *Digraphs: theory, algorithms and applications*. Springer Science & Business Media.
- Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian journal of education*, 33(2), 151-163.
- Bloom, B. Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: McKay Company Inc.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement*, 58(3), 357-381.
- Fink, L. D. (2003). *Creating Significant Learning Experiences*. San Francisco, CA: Jossey-Bass.
- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on. *Educational measurement: issues and practice*, 12(3), 38-47.
- Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Key concepts in educational assessment*. Sage.
- Kean, J., & Reilly, J. (2014). Item response theory. *Handbook for Clinical Research: Design, Statistics and Implementation*.(pp195-198). New York, NY: Demos Medical Publishing.

- Lin, S. Y., & Singh, C. (2013). Can free-response questions be approximated by multiple-choice equivalents?. *American Journal of Physics*, 81(8), 624-629.
- O'Neill, G. (2015). Curriculum Design in Higher Education: Theory to Practice. Retrieved from <http://hdl.handle.net/10197/7137>
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5-13.
- Revelle, W. (2014). psych: Procedures for psychological, psychometric, and personality research. *Northwestern University, Evanston, Illinois*, 165.
- Revelle, W., & French, J. A. (2013). The "New Psychometrics" - Item Response Theory. In Revelle, W., & French, J. (2013). *An introduction to psychometric theory with applications in R*. Retrieved from <http://www.personality-project.org/r/book/#chapter8>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of statistical software*, 17(5), 1-25.
- Scott, M., Stelzer, T., & Gladding, G. (2006). Evaluating multiple-choice exams in large introductory physics courses. *Physical Review Special Topics-Physics Education Research*, 2(2), 020102.
- Sharkness, J. (2014). Item Response Theory: Overview, Applications, and Promise for Institutional Research. *New Directions for Institutional Research*, 2014: 41-58.
- Sharkness, J. (2014). Item Response Theory: Overview, Applications, and Promise for Institutional Research. *New Directions for Institutional Research*, 2014(161), 41-58.
- Tam, M. (2014). Outcomes-based approach to quality assessment and curriculum improvement in higher education. *Quality Assurance in Education*, 22(2), 158-168.
- Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, 29(1), 18.
- Zięba, A. (2013). The item information function in one and two-parameter logistic models—a comparison and use in the analysis of the results of school tests. *Didactics of Mathematics*, (10 (14), 87-96.

\* Ένας κατευθυνόμενος γράφος (digraph)  $D$ , αποτελείται από μη κενό πεπερασμένο σύνολο  $V(D)$  στοιχείων που καλούνται κόμβοι και ένα επίσης πεπερασμένο σύνολο  $A(D)$  διατεταγμένων ζευγών κορυφών που καλούνται τόξα. Συχνά γράφουμε  $D=(V,A)$ , δηλώνοντας το  $V$  ως το σύνολο των κόμβων και  $A$  το σύνολο των τόξων (Bang-Jensen & Gutin, 2008). Η τάξη του γράφου ισούται με το πλήθος των στοιχείων του συνόλου  $A$