ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Σχεδίαση και Ανάπτυξη προσαρμοστικού περιβάλλοντος αξιολόγησης επίδοσης σπουδαστών μέσω διαδικτύου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναστάσιος Χ. Μαυρίδης

Επιβλέπων : Βασίλειος Λούμος
Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2010

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Σχεδίαση και Ανάπτυξη προσαρμοστικού περιβάλλοντος αξιολόγησης επίδοσης σπουδαστών μέσω διαδικτύου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναστάσιος Χ. Μαυρίδης

Επιβλέπων : Βασίλειος Λούμος
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13η Οκτωβρίου 2010.

Αθήνα, Οκτώβριος 2010

.....	.....	.....
Βασίλειος Λούμος	Ελευθέριος Καγιάφας	Μιχαήλ Θεολόγου
Καθηγητής ΕΜΠ	Καθηγητής ΕΜΠ	Καθηγητής ΕΜΠ

Περίληψη

Σκοπός της εργασίας αυτής είναι η μελέτη, η σχεδίαση και η ανάπτυξη ενός προσαρμοστικού διαδικτυακού περιβάλλοντος αξιολόγησης σπουδαστών μέσω ηλεκτρονικού υπολογιστή (Computer Adaptive Testing , C.A.T.). Το C.A.T. χρησιμοποιεί στατιστικά δεδομένα για να παρουσιάσει στον κάθε εξεταζόμενο ένα σύνολο ερωτήσεων πολλαπλών επιλογών. Το σύνολο των ερωτήσεων, σε αντίθεση με την κλασική μέθοδο αξιολόγησης, είναι διαφορετικό για τον κάθε εξεταζόμενο και εξαρτάται από την ικανότητα του. Χρησιμοποιώντας C.A.T. μπορεί να επιτευχθεί μεγαλύτερη ακρίβεια και ασφάλεια στην εκτίμηση της ικανότητας του εξεταζόμενου και ευκολότερη και αποδοτικότερη διαχείριση της κατανομής των ερωτήσεων στα κεφάλαια της εξεταστέας ύλης. Η εκτίμηση της ικανότητας γίνεται αποκλειστικά από τον Η/Υ και δεν απαιτείται διόρθωση διαγωνισμάτων από άνθρωπο. Επιπλέον δίνεται η δυνατότητα στον εξεταζόμενο να δει τα αποτελέσματα αμέσως μετά από την ολοκλήρωση της εξέτασης. Εκτός από τα συνολικά αποτελέσματα υπάρχει η δυνατότητα εμφάνισης αποτελεσμάτων ανά κεφάλαιο της εξεταστέας ύλης για αναλυτικότερη παρουσίαση της απόδοσης του εξεταζόμενου.

Το C.A.T βασίζεται σε στατιστικές μεθόδους για την επιλογή των ερωτήσεων και την εκτίμηση της ικανότητας και συγκεκριμένα στη θεωρία απόκρισης αντικειμένων (Item Response Theory , I.R.T.). Βασικά στοιχεία της θεωρίας αυτής είναι τα στατιστικά μοντέλα, τα οποία εκφράζουν με μαθηματικές σχέσεις τη συμπεριφορά ενός εξεταζόμενου σε ένα διαγώνισμα όσο αφορά τις απαντήσεις του , με βάση τις απαντήσεις που έχει δώσει ένας μεγάλος αριθμός εξεταζόμενων παρόμοιας εκτιμούμενης ικανότητας. Στη μελέτη αυτή παρουσιάζονται οι βασικότεροι αλγόριθμοι και μέθοδοι που χρησιμοποιούνται μέχρι στιγμής στα C.A.T. Χρησιμοποιώντας κάποιες από αυτές τις μεθόδους έχει υλοποιηθεί ένα προσαρμοστικό περιβάλλον αξιολόγησης σπουδαστών.

Λέξεις κλειδιά

Μέθοδος αξιολόγησης, προσαρμοστική, εξέταση, θεωρία απόκρισης αντικειμένων, αξιολόγηση σπουδαστών.

Abstract

The scope of this thesis was the study, design and development of a web based computer adaptive testing (C.A.T.) platform. A C.A.T. uses statistical data to create a multiple choice test for each examinee. Unlike the classic assessment method, the questions of the test are different for each examinee, depending on the examinee ability level. Using C.A.T. for student assessment great measurement precision can be achieved. Easier and efficient content balancing is another advantage. The final score is estimated by computer, so the examinee can get the test results when the test ends. Furthermore we can present more detailed estimations about the examinee's ability such as the ability on each of the course's sections.

C.A.T. selects the test questions and estimates examinee ability based on statistical methods, namely the Item Response Theory (I.R.T.). Basic elements of this theory are the statistical models, which represent the examinee's response behavior in the form mathematical functions, in relation to the responses given by a great amount of examinees with similar ability level. This paper presents the basic algorithms and techniques used in C.A.T. today. Using some of these techniques a C.A.T. platform for student assessment was developed.

Key Words

CAT , computer adaptive test, IRT , item response theory , student assessment

Πίνακας Περιεχόμενων

Α . Γενικά για τις προσαρμοστικές μεθόδους αξιολόγησης (CAT)

Για την αξιολόγηση της επίδοσης ενός ατόμου σε ένα αντικείμενο, είτε σε επίπεδο ψυχολογίας είτε σε επίπεδο εκπαίδευσης, απαιτείται κάποια μαθηματική διατύπωση της επίδοσης αυτής μέσω κάποιας μεταβλητής. Η μεταβλητή αυτή συνήθως περιγράφει μαθηματικά ένα χαρακτηριστικό γνώρισμα της προσωπικότητας του εξεταζόμενου ατόμου, το οποίο δεν είναι άμεσα μετρήσιμο όπως είναι κάποια άλλα χαρακτηριστικά όπως για παράδειγμα το ύψος ή το βάρος του, αλλά μπορούμε να το κατανοήσουμε μόνο διαισθητικά. Όταν κάποιος μας περιγράψει ένα άτομο σαν «καλό στα μαθηματικά» ή «μέτριο στα μαθηματικά» μπορούμε να καταλάβουμε διαισθητικά αυτό που θέλει να μας πει. Παρ' όλα αυτά υπάρχουν περιπτώσεις που χρειάζεται ένα μέγεθος έκφρασης αυτού του χαρακτηριστικού , έτσι ώστε να μπορεί να συγκριθεί ένα άτομο με κάποιο άλλο , ή να έχουμε μια γενική ιδέα σχετικά με το ποσοστό του χαρακτηριστικού που κατέχει το άτομο σε σχέση με το ποσοστό που κατέχει το μέσο άτομο.

Παρόμοια μπορεί κάποιος να μιλήσει για την ικανότητα ενός ατόμου ως σπουδαστή και τις ιδιότητες της, όπως το πόσο καλούς βαθμούς έχει το άτομο αυτό , το πόσο εύκολα μπορεί να κατανοήσει κάποιο νέο αντικείμενο, το πόσο δημιουργικά χρησιμοποιεί το χρόνο που έχει για να μελετήσει κτλ. Καθεμία από τις παραπάνω ιδιότητες μπορούν να περιγραφούν με κάποια μεταβλητή. Αν και μια τέτοια μεταβλητή μπορεί να οριστεί θεωρητικά και να περιγραφεί εύκολα, δεν μπορεί να μετρηθεί άμεσα γιατί δεν εκφράζει κάποιο φυσικό μέγεθος αλλά κάποια αφηρημένη έννοια. Ο κύριος σκοπός των εκπαιδευτικών και ψυχολογικών μετρήσεων είναι να καθορίσουν το πόσο από ένα ή περισσότερα χαρακτηριστικά γνωρίσματα κατέχει ένα άτομο. Στη βιβλιογραφία της Θεωρίας Απόκρισης Αντικειμένων χρησιμοποιείται ο γενικός όρος «**ικανότητα**» για αναφορές σε τέτοια χαρακτηριστικά γνωρίσματα . Αυτός ο όρος θα χρησιμοποιηθεί και στην παρούσα μελέτη.

Τις περισσότερες φορές η εκτίμηση της ικανότητας ενός ατόμου γίνεται μέσω κάποιας εξέτασης, είτε γραπτής είτε μέσω Η/Υ, στην οποία οι εξεταζόμενοι καλούνται να απαντήσουν σε κοινές ερωτήσεις. Όλοι οι εξεταζόμενοι καλούνται να απαντήσουν στο ίδιο σταθερό σύνολο ερωτήσεων, έτσι ο κάθε εξεταζόμενος πολλές φορές καλείται να απαντήσει σε ερωτήσεις πολύ δύσκολες ή πολύ εύκολες για αυτόν. Οι ερωτήσεις αυτές παρέχουν πολύ μικρή πληροφορία για το επίπεδο της ικανότητας του, και έτσι χρειάζεται μεγάλος αριθμός ερωτήσεων για να υπολογιστεί με ακρίβεια η ικανότητα αυτή.

Ο σκοπός μιας προσαρμοστικής μεθόδου αξιολόγησης (Computer Adaptive Testing) είναι να δημιουργήσει το βέλτιστο ερωτηματολόγιο για κάθε εξεταζόμενο χρησιμοποιώντας κάποιες από τις ερωτήσεις που έχουν τοποθετηθεί σε μια βάση δεδομένων. Για να επιτευχθεί αυτό γίνεται εκτίμηση του προς μέτρηση χαρακτηριστικού της προσωπικότητας του εξεταζόμενου κατά τη διάρκεια της εξέτασης και επιλέγονται από τη βάση δεδομένων τις ερωτήσεις που είναι κατάλληλες για την εκτιμούμενη ικανότητα (θ) του. Οι ερωτήσεις επιλέγονται έτσι ώστε να ταιριάζουν με την ικανότητα θ του εξεταζόμενου με βάση στατιστικές μεθόδους και συγκεκριμένα τη θεωρία απόκρισης αντικειμένων (Item Response Theory , I.R.T.). Βασικά στοιχεία της θεωρίας αυτής είναι τα στατιστικά μοντέλα , τα οποία εκφράζουν με μαθηματικές σχέσεις τη συμπεριφορά απόκρισης ενός εξεταζόμενου σε μια

εξέταση, με βάση τις απαντήσεις που έχει δώσει στις ίδιες ερωτήσεις ένας μεγάλος αριθμός εξεταζόμενων παρόμοιας εκτιμούμενης ικανότητας. Η ικανότητα εκτιμάται με επαναληπτικές μεθόδους σε κάθε βήμα της εξέτασης. Η εκτίμηση δίνεται σε σχέση με την ικανότητα ενός κανονικού δείγματος εξεταζόμενων , δηλ. ενός δείγματος εξεταζόμενων των οποίων οι ικανότητες ακολουθούν κανονική κατανομή. Είναι φανερό λοιπόν ότι για την εφαρμογή της I.R.T. είναι απαραίτητη η ύπαρξη μιας βάσης δεδομένων , η οποία περιέχει στατιστικά στοιχεία για κάθε ερώτηση που θα χρησιμοποιηθεί στην εξέταση. Για να δημιουργηθεί μια τέτοια βάση απαιτείται η διεξαγωγή κάποιων δοκιμαστικών εξετάσεων, στις οποίες ένας μεγάλος αριθμός εξεταζόμενων καλείται να απαντήσει στις ίδιες ερωτήσεις. Στην ορολογία της ψυχομετρίας και της I.R.T. οι ερωτήσεις που καλείται να απαντήσει ένας εξεταζόμενος ονομάζονται **αντικείμενα**. Αυτός ο όρος θα χρησιμοποιηθεί και στην παρούσα μελέτη. Τα βασικά στοιχεία της I.R.T. αναλύονται στο κεφάλαιο Β της μελέτης.

Σε αντίθεση με την κλασική μέθοδο , σε μια προσαρμοστική εξέταση C.A.T. διαφορετικοί εξεταζόμενοι καλούνται να απαντήσουν σε διαφορετικά σύνολα ερωτήσεων και σε διαφορετικό αριθμό ερωτήσεων. Επιπλέον ο εξεταζόμενος απαντά σε ερωτήσεις που είναι κατάλληλες για το επίπεδο ικανότητάς του και έτσι η ικανότητά του μπορεί να εκτιμηθεί με μεγαλύτερη ακρίβεια , χωρίς να χρειάζεται μεγάλος αριθμός ερωτήσεων , όπως στην κλασική μέθοδο.

Η αρχική έρευνα για το CAT έγινε τη δεκαετία του 70 και του 80 και συνεχίζεται μέχρι και σήμερα. Τα ερευνητικά πεδία που σχετίζονται με το CAT σήμερα είναι η ανάπτυξη βάσεων αντικειμένων, βελτιστοποίηση των διαδικασιών επιλογής αντικειμένου, εναλλακτικές μέθοδοι για την εκτίμηση των χαρακτηριστικών γνωρισμάτων του εξεταζόμενου, η ασφάλεια των διαγωνισμάτων και η αξιοπιστία των αποτελεσμάτων και άλλα σχετικά θέματα.

Μερικές από τις εξετάσεις που γίνονται με μεθόδους C.A.T. είναι το GRE (Graduate Record Examination), το GMAT (Graduate Management Admissions Test), το ASVAB (Armed Services Vocational Aptitude Test Battery) όπως και πολλά από τα διαγωνίσματα της εταιρίας Microsoft για τα πτυχία Microsoft Certified Professional και τα διαγωνίσματα για τα πτυχία πιστοποιήσεως γνώσεων από την CompTIA (Computing Technology Industry Association). Επίσης μέθοδοι C.A.T. χρησιμοποιούνται στα διαγωνίσματα που έχουν αναπτυχθεί από το Εθνικό Ινστιτούτο Εκπαιδευτικών Μετρήσεων της Ολλανδίας (National Institute for Educational Measurement of Netherlands), στο **LPCAT** που αναπτύχθηκε στο Πανεπιστήμιο της Νοτίου Αφρικής , και σε πολλές άλλες εξετάσεις ανά τον κόσμο.

Μέθοδοι C.A.T. χρησιμοποιήθηκαν αρχικά στην ψυχολογία και την ψυχομετρία με τη μορφή διαφόρων ειδών τεστ προσωπικότητας, και αργότερα βρήκαν εφαρμογή σε εκπαιδευτικά λογισμικά για την αξιολόγηση της επίδοσης των εκπαιδευόμενων. Άλλες δημοφιλείς εφαρμογές περιλαμβάνουν ιατρικά λογισμικά που σχετίζονται με την αξιολόγηση της κατάστασης κάποιου ασθενούς αλλά και την πρόγνωση ασθενειών.

Ένα C.A.T. έχει πολλά πλεονεκτήματα σε σχέση με την κλασική μέθοδο εξετάσεων με γραπτούς διαγωνισμούς. Τα μικρότερα σε μέγεθος διαγωνίσματα , η μεγαλύτερη ακρίβεια στη βαθμολογία, η εξέταση on demand, η άμεση βαθμολόγηση και τα άμεσα αναλυτικά αποτελέσματα είναι τα κυριότερα από αυτά.

Από την άλλη πλευρά το κόστος είναι αρκετά υψηλότερο, αν αναλογιστούμε τους ανθρώπους που πρέπει να δουλέψουν για να υλοποιήσουν και να οργανώσουν μια τέτοια εξέταση. Ένα άλλο πρόβλημα είναι η ασφάλεια της εξέτασης, η οποία ενώ αρχικά ήταν ένα από τα μεγαλύτερα πλεονεκτήματα του C.A.T., έχει γίνει ένα από τα σημαντικότερα προβλήματά του. Για να εξασφαλίσουμε την ασφάλεια και την αξιοπιστία των αποτελεσμάτων, οι ερωτήσεις που βρίσκονται στη βάση δεδομένων του προγράμματος πρέπει να ανανεώνονται συνεχώς, και αυτό έχει αυξήσει σημαντικά το κόστος υλοποίησης. Παρ' όλα τα προβλήματά τους όμως τα πλεονεκτήματα των εφαρμογών που χρησιμοποιούν C.A.T. υποσκελίζουν τα μειονεκτήματά τους.

Β . Η θεωρία απόκρισης αντικειμένων (IRT , Item Response Theory)

Ο D.N. Lawley δημοσίευσε το 1943 μια μελέτη στην οποία έδειχνε ότι πολλά από τα δομικά χαρακτηριστικά της κλασικής θεωρίας εξετάσεων μπορούσαν να εκφραστούν σαν παράμετροι της χαρακτηριστικής γραφικής παράστασης ενός αντικειμένου. Αυτή η μελέτη σηματοδοτεί την αρχή της Θεωρίας Απόκρισης Αντικειμένων σαν θεωρία για τη μέτρηση κάποιου χαρακτηριστικού. Η θεωρία αυτή αναπτύχθηκε περεταίρω, οι ορισμοί της διατυπώθηκαν σαφέστερα και συστηματικότερα, και αργότερα επεκτάθηκε. Τις προηγούμενες δεκαετίες αναπτύχθηκε το απαραίτητο λογισμικό για την εφαρμογή της και στην πράξη.

Η IRT αποτελείται από ένα σύνολο μαθηματικών συναρτήσεων , οι οποίες μοντελοποιούν την αλληλεπίδραση μεταξύ ενός ατόμου που απαντά (συνήθως μέσω υπολογιστή) σε κάποια αντικείμενα που του δίνονται. Τα μοντέλα, που χρησιμοποιεί η θεωρία για το σκοπό αυτό, προσπαθούν να προβλέψουν την πιθανότητα το άτομο να δώσει μια απάντηση που ανήκει σε μια κατηγορία. Οι κατηγορίες που χρησιμοποιούνται πιο συχνά είναι η «σωστό» και η «λάθος». Προσπαθούν δηλαδή κατά κάποιο τρόπο να προβλέψουν την επίδοση του ατόμου σε κάθε αντικείμενο. Για να το πετύχουν αυτό χρησιμοποιούν κάποια χαρακτηριστικά των εξεταζόμενων και των αντικειμένων, τα οποία αντιστοιχούν σε κάποιες παραμέτρους των συναρτήσεων της θεωρίας. Τα χαρακτηριστικά των εξεταζόμενων σχετίζονται με κάποια μετρήσιμα γνωρίσματά τους, όπως το επίπεδο της ικανότητας σε κάποιο τομέα, το ποσοστό ικανοποίησής του από κάποια γεγονότα ή το ποσοστό συμφωνίας του με κάποια άποψη. Τα χαρακτηριστικά των αντικειμένων αντιστοιχούν σε παραμέτρους που χρησιμοποιούνται από τη θεωρία για να περιγράψουν τη σχέση του αντικειμένου με το προς εκτίμηση χαρακτηριστικό γνώρισμα.

Η Θεωρία Απόκρισης Αντικειμένων, η οποία βασίζεται σε αντικείμενα και όχι βαθμολογίες διαγωνισμάτων για την εκτίμηση της ικανότητας, είναι αρκετά ισχυρότερη από την κλασική

μέθοδο βαθμολόγησης. Αν και οι βασικές έννοιες της θεωρίας αυτής είναι απλές, οι μαθηματικοί υπολογισμοί που απαιτούνται είναι περισσότεροι και υψηλότερου επιπέδου από αυτούς που απαιτούνται στην κλασική θεωρία, και οι αλγόριθμοι για τον υπολογισμό τους αρκετά απαιτητικοί όσο αφορά την υπολογιστική ισχύ. Για το λόγο αυτό μόνο την τελευταία δεκαετία, με τη ραγδαία εξέλιξη στο χώρο της πληροφορικής, δόθηκε η δυνατότητα για την πλήρη εφαρμογή των αλγορίθμων αυτών σε πρακτικό επίπεδο.

Στην IRT, για τη μέτρηση της ικανότητας, χρησιμοποιείται μια κλίμακα με κέντρο το μηδέν, που εκτείνεται από $-\infty$ ως $+\infty$. Η ιδέα πίσω από αυτή την κλίμακα είναι ότι αν μπορούσαμε με κάποιο τρόπο να μετρήσουμε την ικανότητα ενός ατόμου σαν φυσικό μέγεθος, μπορούμε με αυτή την κλίμακα να συγκρίνουμε τις ικανότητες δύο ή περισσότερων ατόμων. Χρησιμοποιείται δηλαδή σαν κλίμακα σύγκρισης. Ενώ η θεωρητικά η ικανότητα παίρνει τιμές από $-\infty$ ως $+\infty$ η πρακτική εφαρμογή της θεωρίας επιβάλλει τον περιορισμό αυτού του πεδίου τιμών, επειδή υπάρχει περιορισμός στο μέγιστο και ελάχιστο αριθμό που μπορεί να αναπαρασταθεί από έναν υπολογιστή και λόγω του περιορισμένου αριθμού ερωτήσεων και εξεταζόμενων. Συνήθως χρησιμοποιείται μια κλίμακα από -3 ως +3. Αυτή η κλίμακα χρησιμοποιήθηκε και στη μελέτη αυτή.

Η συνήθης τακτική που ακολουθείται για τον υπολογισμό της ικανότητας είναι μέσω μιας εξέτασης στην οποία οι εξεταζόμενοι καλούνται να απαντήσουν σε ένα σύνολο αντικειμένων (ερωτήσεων). Καθένα από αυτά τα αντικείμενα για κάθε εξεταζόμενο καθορίζει ένα μέρος της ικανότητας του. Θεωρητικά, ο εξεταζόμενος έπρεπε να μπορεί να δώσει οποιαδήποτε απάντηση αυτός ήθελε στα αντικείμενα και ένας διορθωτής να αποφασίζει αν η απάντηση είναι σωστή ή όχι και να βαθμολογεί ανάλογα το συγκεκριμένο αντικείμενο.

Σύμφωνα με την κλασική θεωρία βαθμολόγησης ο τελικός βαθμός του εξεταζόμενου θα ήταν το άθροισμα των βαθμολογιών σε όλα τα αντικείμενα της εξέτασης. Αυτή η βαθμολογία ονομάζεται ακατέργαστη βαθμολογία (raw test score). Αυτό που μας ενδιαφέρει στην IRT είναι το αν ο εξεταζόμενος απάντησε σωστά ή λάθος καθένα αντικείμενο ξεχωριστά και όχι το σύνολο των σωστών απαντήσεων που έχει δώσει δηλ. το raw test score του. Αυτό συμβαίνει γιατί η IRT βασίζεται στα αντικείμενα ξεχωριστά. Αν δηλαδή ένας εξεταζόμενος κληθεί να απαντήσει σε ένα διαγώνισμα με ερωτήσεις πολύ δύσκολες για αυτόν , σε ένα άλλο με ερωτήσεις πολύ εύκολες για αυτόν και σε ένα τρίτο με ερωτήσεις που απευθύνονται στο επίπεδο ικανότητάς του, εφαρμόζοντας και στις τρεις περιπτώσεις την IRT η εκτιμούμενη ικανότητα του εξεταζόμενου θα κυμαίνεται στα ίδια επίπεδα και κοντά στην πραγματική ικανότητά του. Με την κλασική θεωρία βαθμολόγησης η τελικές εκτιμούμενες ικανότητες στις τρεις περιπτώσεις θα είχαν μεγάλες διαφορές.

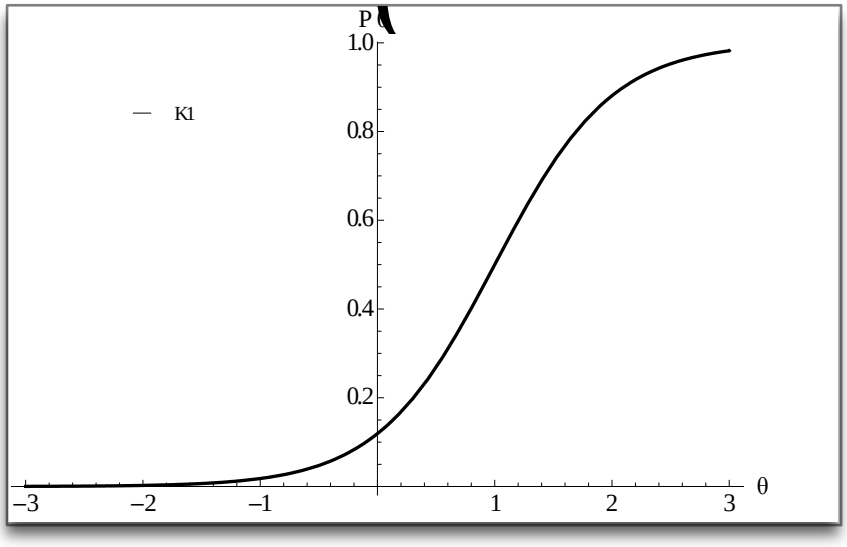
Στην πράξη αντικείμενα που μπορούν να απαντηθούν ελεύθερα είναι δύσκολο να χρησιμοποιηθούν σε μια εξέταση μέσω Η/Υ. Ακόμα και αν η εξέταση δεν γίνεται μέσω Η/Υ τέτοια αντικείμενα είναι δύσκολο να βαθμολογηθούν με ακρίβεια. Επιπλέον όταν η απάντηση βαθμολογείται από έναν υπολογιστή, τέτοια αντικείμενα εισάγουν επιπλέον δυσκολίες στην υλοποίηση και μεγαλύτερο σφάλμα στη βαθμολόγηση. Έτσι τα

περισσότερα διαγωνίσματα που γίνονται χρησιμοποιώντας την IRT περιλαμβάνουν ερωτήσεις πολλαπλών επιλογών.

Κάθε εξεταζόμενος ο οποίος καλείται να συμπληρώσει ένα διαγώνισμα κατέχει ένα ποσό της προς μέτρηση ικανότητας. Έτσι, μπορούμε να χρησιμοποιήσουμε μια αριθμητική τιμή, δηλαδή ένα βαθμό, ο οποίος τοποθετεί τον εξεταζόμενο σε ένα σημείο της κλίμακας ικανοτήτων. Ο βαθμός της ικανότητας συμβολίζεται στη σχετική βιβλιογραφία με το ελληνικό γράμμα θ. Η πιθανότητα κάποιος εξεταζόμενος με βαθμό ικανότητας θ να απαντήσει σωστά σε ένα συγκεκριμένο αντικείμενο συμβολίζεται με P(θ). Σε ένα τυπικό αντικείμενο ενός διαγωνίσματος, αυτή η πιθανότητα θα είναι μικρή για εξεταζόμενους μικρής ικανότητας και μεγάλη για εξεταζόμενους μεγαλύτερης ικανότητας. Προφανώς η πιθανότητα να δώσει ο εξεταζόμενος κάποια λανθασμένη απάντηση είναι Q(θ)=1-P(θ).

Β.1 Η χαρακτηριστική καμπύλη αντικειμένου (Item Characteristic Curve)

Η χαρακτηριστική καμπύλη του αντικειμένου είναι η γραφική αναπαράσταση της πιθανότητας P(θ) συναρτήσει της ικανότητας θ και έχει τη μορφή S. Στο Σχήμα 1 φαίνεται μια τυπική χαρακτηριστική αντικειμένου. Όπως φαίνεται και στο σχήμα η πιθανότητα σωστής απάντησης είναι κοντά στο μηδέν για τα χαμηλότερα επίπεδα ικανότητας, αυξάνει καθώς αυξάνει και το επίπεδο ικανότητας και πλησιάζει το 1 για τα υψηλότερα επίπεδα. Κάθε αντικείμενο ενός διαγωνίσματος έχει τη δική του χαρακτηριστική καμπύλη. Η χαρακτηριστική καμπύλη αντικειμένου αποτελεί το σημαντικότερο στοιχείο της IRT.



Σχήμα 1 : *Η χαρακτηριστική καμπύλη αντικειμένου για ένα τυπικό αντικείμενο.*

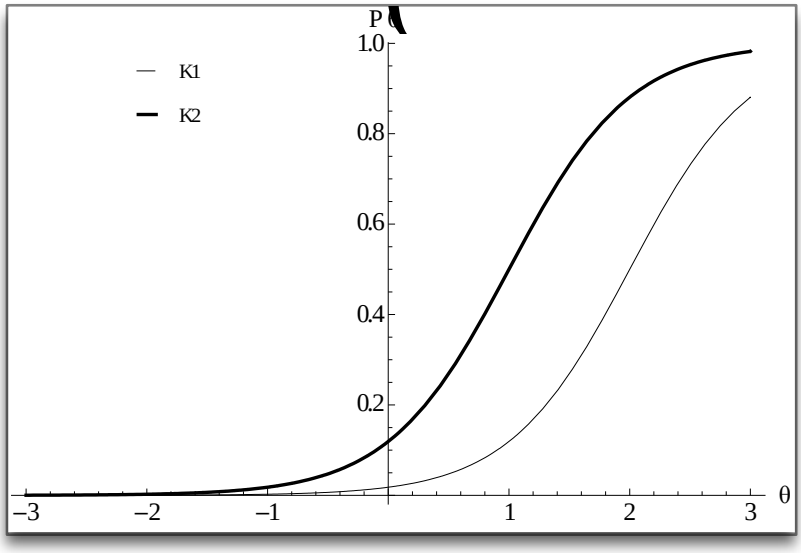
Υπάρχουν δύο βασικά τεχνικά χαρακτηριστικά της καμπύλης , τα οποία χρησιμοποιούνται για να την περιγράψουν. Το πρώτο είναι η **δυσκολία** του αντικειμένου. Σύμφωνα με τη θεωρία, η δυσκολία περιγράφει την περιοχή στην κλίμακα ικανοτήτων στην οποία το αντικείμενο είναι αποτελεσματικό. Για παράδειγμα ένα εύκολο αντικείμενο είναι αποτελεσματικό για εξεταζόμενους χαμηλού επιπέδου ικανότητας, ενώ ένα δύσκολο αντικείμενο για εξεταζόμενους υψηλού επιπέδου ικανότητας. Έτσι κάθε αντικείμενο μπορεί

να τοποθετηθεί σύμφωνα με τη δυσκολία του στον άξονα των ικανοτήτων. Το δεύτερο χαρακτηριστικό είναι ο **παράγοντας διακριτικής ικανότητας**, ο οποίος εκφράζει το πόσο καλά μπορεί το αντικείμενο να διακρίνει εξεταζόμενους με ικανότητα μικρότερη από τη θέση του αντικειμένου στην κλίμακα των ικανοτήτων από εξεταζόμενους με μεγαλύτερη από αυτή τη θέση ικανότητα. Διαισθητικά ο παράγοντας αυτός αντιπροσωπεύει την κλίση της καμπύλης στο μέσο της. Όσο πιο απότομη είναι η κλίση αυτή τόσο καλύτερα μπορεί το αντικείμενο να διαχωρίσει τους εξεταζόμενους. Όσο η κλίση μειώνεται τόσο δυσκολότερα μπορεί το αντικείμενο να διαχωρίσει τους εξεταζόμενους γιατί η πιθανότητα σωστής απάντησης για χαμηλά επίπεδα ικανότητας είναι περίπου ίση με την πιθανότητα για υψηλότερα επίπεδα ικανότητας. Αυτοί οι δύο παράγοντες της καμπύλης δεν μας δίνουν καμία πληροφορία για το αν το συγκεκριμένο αντικείμενο πράγματι μετρά κάποιο μέρος της ικανότητας του εξεταζόμενου. Χρησιμοποιούνται για να περιγράψουν τη γενική μορφή της χαρακτηριστικής καμπύλης του αντικειμένου.

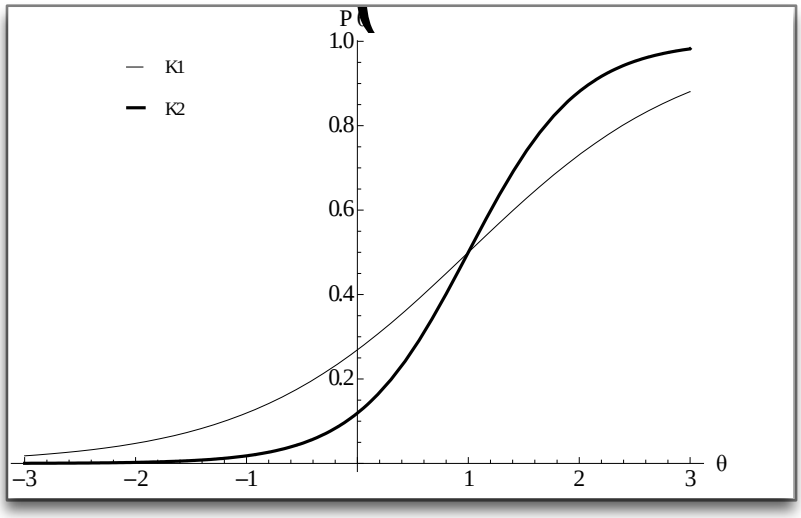
Στο Σχήμα 2 φαίνονται οι χαρακτηριστικές δύο αντικειμένων I1 , I2. Και τα δύο έχουν ίδιο παράγοντα διάκρισης αλλά διαφορετική δυσκολία. Οι χαρακτηριστική που βρίσκεται αριστερά (I1) αντιστοιχεί σε ευκολότερο αντικείμενο, καθώς η πιθανότητα σωστής απάντησης είναι υψηλή για χαμηλού επιπέδου ικανότητας εξεταζόμενους και πλησιάζει το 1 για υψηλότερου επιπέδου εξεταζόμενους. Παρατηρούμε επίσης ότι για αντικείμενα μεγάλης δυσκολίας (όπως το I2) η πιθανότητα σωστής απάντησης είναι μικρότερη του 1 ακόμα και για εξεταζόμενους ικανότητας 3. Ανάλογα λοιπόν με τη δυσκολία ενός αντικειμένου η καμπύλη μετακινείται αριστερά ή δεξιά στον άξονα των ικανοτήτων.

Στο Σχήμα 3 φαίνονται οι χαρακτηριστικές καμπύλες δύο αντικειμένων I1 και I2 που έχουν την ίδια δυσκολία αλλά διαφορετικούς παράγοντες διάκρισης. Το αντικείμενο I2 του οποίου η χαρακτηριστική έχει μεγαλύτερη κλίση έχει μεγαλύτερη διακριτική ικανότητα από ότι το I1. Στην περιοχή γύρω από το σημείο θ=1 και καθώς το θ αυξάνει η πιθανότητα σωστής απάντησης του αντικειμένου I2 αυξάνει με μεγαλύτερο ρυθμό από ότι αυτή του I1.

Ένα αντικείμενο με ιδανικό παράγοντα διάκρισης θα έχει τέλεια διακριτική ικανότητα. Η χαρακτηριστική καμπύλη ενός τέτοιου αντικειμένου θα ήταν μηδέν μέχρι το σημείο $\theta=\theta_1$, όπου θ_1 η δυσκολία του αντικειμένου , και ένα από $\theta=\theta_1$ ως θ=3. Ένα τέτοιο αντικείμενο όμως δε θα μπορούσε να διαχωρίσει καθόλου δυο εξεταζόμενους που έχουν και οι δύο ικανότητα μικρότερη της θ_1 , ούτε ασφαλώς δύο άλλους που έχουν και οι δύο ικανότητα μεγαλύτερη του θ_1



Σχήμα 2 : Οι χαρακτηριστικές καμπύλες αντικειμένου για δύο αντικείμενα με ίδιους παράγοντες διακριτικής ικανότητας αλλά διαφορετικών δυσκολιών.



Σχήμα 3: Οι χαρακτηριστικές καμπύλες αντικειμένου για δύο αντικείμενα ίδιας δυσκολίας αλλά διαφορετικών παραγόντων διακριτικής ικανότητας

Β.1.1 Μοντέλα χαρακτηριστικών καμπυλών αντικειμένου

Για να μελετήσουμε επιστημονικά τα αντικείμενα της βάσης ενός CAT, η ορολογία που χρησιμοποιήθηκε ως τώρα δεν είναι αρκετή. Περιγράφοντας ένα αντικείμενο σαν «δύσκολο» , «εύκολο» ή ότι έχει «καλή διακριτική ικανότητα» μας δίνει μια γενική ιδέα για τη μορφή της χαρακτηριστικής καμπύλης του, αλλά δεν μας παρέχει τη ζητούμενη ακρίβεια. Στην IRT έχουν αναπτυχθεί αρκετά μοντέλα καμπυλών αντικειμένου, το καθένα από τα οποία παρέχει μια μαθηματική έκφραση της πιθανότητας σωστής απάντησης συναρτήσει της ικανότητας. Κάθε μοντέλο εισάγει μια ή περισσότερες παραμέτρους. Οι τιμές αυτών των παραμέτρων καθορίζουν μια συγκεκριμένη χαρακτηριστική καμπύλη αντικειμένου. Η εισαγωγή των παραμέτρων και κατά συνέπεια η χρήση των μοντέλων γίνεται για να μπορέσουμε να εφαρμόσουμε κάποια θεωρία μέτρησης η οποία να είναι καλά ορισμένη και να μας παρέχει την απαιτούμενη ακρίβεια.

Η βαθμολόγηση των αντικειμένων μπορεί να γίνει με δύο τρόπους, είτε διχότομα (dichotomous) είτε πολυχότομα(polychotomous). Όταν χρησιμοποιούνται διχότομα αντικείμενα, η σωστή απάντηση βαθμολογείται με 1 και καθεμία από τις λανθασμένες με 0.

Όταν χρησιμοποιούνται πολυχότομα αντικείμενα οι εναλλακτικές απαντήσεις κάθε αντικειμένου χωρίζονται σε κατηγορίες και η κάθε κατηγορία βαθμολογείται με ένα μέρος του συνολικού βαθμού που αντιστοιχεί στο συγκεκριμένο αντικείμενο. Υπάρχουν μοντέλα για την περίπτωση που χρησιμοποιούνται διχότομα βαθμολογημένα αντικείμενα όσο και για αυτή που χρησιμοποιούνται πολυχότομα βαθμολογημένα. Τα περισσότερα μοντέλα για πολυχότομα αντικείμενα προκύπτουν από τη γενίκευση ή την παραλλαγή αυτών για διχότομα. Στην περίπτωση που χρησιμοποιούνται πολυχότομα αντικείμενα, με κατάλληλη επιλογή των παραμέτρων, το μοντέλο που χρησιμοποιούμε μπορεί να συμπέσει με ένα από τα μοντέλα για διχότομα. Έτσι μπορούμε να χρησιμοποιήσουμε ένα μοντέλο για πολυχότομα αντικείμενα, έστω και αν η βάση μας αποτελείται από διχότομα, με κατάλληλη επιλογή των παραμέτρων.

Β.1.2 Η Συνάρτηση Logistic

Το σύνηθες μοντέλο για τη χαρακτηριστική καμπύλη αντικειμένου στην IRT είναι αυτό της συνάρτησης Logistic. Περιγράφει μια οικογένεια καμπυλών οι οποίες έχουν γενικά το σχήμα S, όπως αυτή στο Σχήμα 1. Η συνάρτηση αυτή χρησιμοποιήθηκε για πρώτη φορά σαν μοντέλο της χαρακτηριστικής καμπύλης αντικειμένου τη δεκαετία του 50 και, λόγω της απλότητάς της, προτιμάται από τους επιστήμονες ως τις μέρες μας. Ο μαθηματικός τύπος της συνάρτησης Logistic είναι

$$P(\theta) = \frac{1}{1 + e^{-L}} \qquad (\Sigma.1)$$

Όπου το L ονομάζεται logit της συνάρτησης και είναι συνάρτηση του θ. Ο γενικός τύπος για το L είναι L=-a(θ-b), αλλά διαφέρει ανάλογα με το μοντέλο που χρησιμοποιούμε.
Το b ονομάζεται **παράγοντας δυσκολίας** αντικειμένου και είναι το σημείο στον άξονα των ικανοτήτων στο οποίο η πιθανότητα σωστής απάντησης για το αντικείμενο αυτό είναι 0,5 σε πολλά μοντέλα.
Το πεδίο τιμών του παράγοντα δυσκολίας είναι θεωρητικά [-∞ , + ∞] αλλά πρακτικά οι τυπικές τιμές του είναι στο διάστημα [-3 , +3].

Το a ονομάζεται **παράγοντας διακριτικής ικανότητας αντικειμένου**. Επειδή η χαρακτηριστική καμπύλη αντικειμένου έχει σχήμα S, η κλίση της καμπύλης αλλάζει συναρτήσει της ικανότητας και παίρνει τη μέγιστη τιμή της στο σημείο όπου το επίπεδο ικανότητας ισούται με τη δυσκολία του αντικειμένου ($\hat{\theta} = b$). Ο παράγοντας διακριτικής ικανότητας του αντικειμένου είναι ανάλογος με την κλίση της καμπύλης στο σημείο $\hat{\theta} = b$.
Στα συνήθη μοντέλα η κλίση της καμπύλης στο σημείο αυτό είναι a/4. Το πεδίο τιμών του παράγοντα αυτού είναι θεωρητικά [-∞ , + ∞] αλλά πρακτικά οι τυπικές τιμές του είναι στο διάστημα [-2.8 , +2.8].

Επειδή στο σημείο $\hat{\theta} = b$ η κλίση της χαρακτηριστικής καμπύλης αντικειμένου είναι μέγιστη, η παράμετρος δυσκολίας ορίζει το σημείο του άξονα των ικανοτήτων για το οποίο το αντικείμενο λειτουργεί καλύτερα. Ορίζει δηλαδή την ικανότητα των εξεταζόμενων για την οποία η ερώτηση αυτή είναι καταλληλότερη.

Β.1.3 Μοντέλα χαρακτηριστικών καμπυλών αντικειμένου για διχότομα βαθμολογημένα αντικείμενα

Β.1.3.1 Το Μοντέλο μιας παραμέτρου ή Μοντέλο Rasch (Rasch model ή One Parameter Logistic Model ή 1PL model)

Ο Δανός μαθηματικός Georg Rasch εισήγαγε τη δεκαετία του 60 το μοντέλο μιας παραμέτρου για τη χαρακτηριστική καμπύλη αντικειμένου. Για την ανάλυσή των πειραματικών δεδομένων χρησιμοποίησε μεθόδους βασισμένες στη θεωρία των πιθανοτήτων. Το αποτέλεσμα της ανάλυσής του ήταν ένα μοντέλο που βασίζεται στη συνάρτηση Logistic. Σε αυτό το μοντέλο ο παράγοντας διακριτικής ικανότητας του αντικειμένου είναι σταθερός και ίσος με ένα (a=1) για όλα τα αντικείμενα. Μεταβλητός είναι μόνο ο παράγοντας δυσκολίας b, που είναι διαφορετικός για κάθε αντικείμενο.

Η πιθανότητα σωστής απάντησης ενός αντικειμένου στο μοντέλο αυτό δίνεται από τον τύπο

$$P(\theta) = \frac{1}{1 + e^{-1(\theta - b)}} \qquad (\Sigma. 2)$$

Όπου b ο παράγοντας δυσκολίας του αντικειμένου και θ η ικανότητα.
Το b αντιστοιχεί στο σημείο του άξονα των ικανοτήτων στο οποίο η πιθανότητα σωστής απάντησης για το αντικείμενο αυτό είναι 0,5.

Επειδή ο παράγοντας διακριτικής ικανότητας στο μοντέλο αυτό είναι σταθερός και ίσος με a=1 για όλα τα αντικείμενα, η κλίση της χαρακτηριστικής καμπύλης αντικειμένου στο σημείο θ=b είναι σταθερή και ίση με 1/4 για όλα τα αντικείμενα. Αυτό που αλλάζει στις χαρακτηριστικές διαφορετικών αντικειμένων είναι η θέση στην οποία είναι θ=b, δηλαδή η θέση στην οποία ισχύει P(θ)=0,5.

Β.1.3.2 Το μοντέλο δύο παραμέτρων (Two Parameter Logistic Model ή 2PL)

Η πιθανότητα σωστής απάντησης ενός αντικειμένου στο μοντέλο αυτό δίνεται από τον τύπο

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}} \quad (\Sigma. 3)$$

Όπου b ο παράγοντας δυσκολίας του αντικειμένου, a ο παράγοντας διακριτικής ικανότητας και θ η ικανότητα. Και σε αυτό το μοντέλο το b αντιστοιχεί στο σημείο του άξονα των ικανοτήτων στο οποίο η πιθανότητα σωστής απάντησης για το αντικείμενο αυτό είναι 0,5.

Σε αυτό το μοντέλο ο παράγοντας δυσκολίας b και ο παράγοντας διακριτικής ικανότητας είναι διαφορετικοί για κάθε αντικείμενο.

Το μοντέλο αυτό αναφέρεται συχνά στη βιβλιογραφία σαν μοντέλο Birnbaum, επειδή ο Birnbaum ήταν αυτός που το εισήγαγε. Ο ίδιος βέβαια έχει εισάγει και άλλα μοντέλα, όπως είναι το μοντέλο τριών παραμέτρων που αναλύεται στην επόμενη παράγραφο, και αυτό αρκετές φορές δημιουργεί σύγχυση. Γι αυτό το λόγο το μοντέλο αυτό θα αναφέρεται σαν μοντέλο δύο παραμέτρων ή 2PL στη μελέτη αυτή.

Β.1.3.3 Το μοντέλο τριών παραμέτρων (Three Parameter Logistic Model ή 3PL)

Κανένα από τα δύο μοντέλα που έχουν αναφερθεί παραπάνω δεν λαμβάνει υπόψη τον παράγοντα της τύχης στον υπολογισμό της πιθανότητας σωστής απάντησης ενός αντικειμένου. Ένας παράγοντας της πιθανότητας αυτής όμως είναι σίγουρα και η τύχη καθώς κάποιος μπορεί να μαντέψει τη σωστή απάντηση και έτσι η βαθμολογία του να μη βασίζεται αποκλειστικά στην ικανότητά του. Ο Birnbaum εισήγαγε το 1968 ένα μοντέλο το οποίο περιλάμβανε έναν παράγοντα που εκφράζει τη συνεισφορά της τύχης στον προσδιορισμό της πιθανότητας σωστής απάντησης. Το μοντέλο αυτό προήρθε από την τροποποίηση του μοντέλου δύο παραμέτρων. Αν και με την τροποποίηση αυτή το μοντέλο από τεχνική άποψη δεν ανήκει στην κατηγορία των Logistic είναι γνωστό σαν three parameter logistic model.

Η πιθανότητα σωστής απάντησης ενός αντικειμένου στο μοντέλο αυτό δίνεται από τον τύπο

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-a(\theta - b)}} \quad (\Sigma.4)$$

Όπου b ο παράγοντας δυσκολίας του αντικειμένου , a ο παράγοντας διακριτικής ικανότητας, c ο παράγοντας τύχης και θ η ικανότητα.

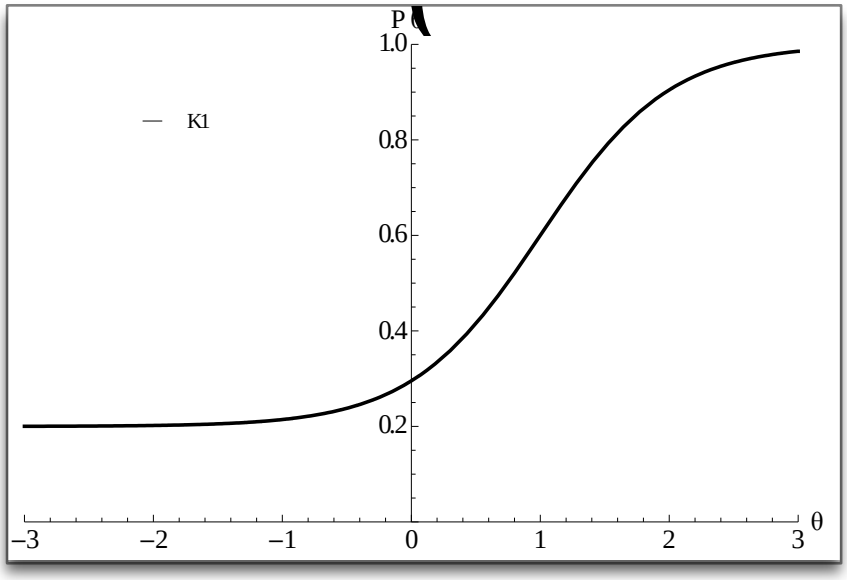
Ο παράγοντας τύχης c εκφράζει τη συνεισφορά της τύχης στον προσδιορισμό της πιθανότητας σωστής απάντησης και ισούται με την πιθανότητα κάποιος εξεταζόμενος να απαντήσει σωστά στο αντικείμενο αυτό μαντεύοντας τη σωστή απάντηση. Εξ ορισμού αυτός ο παράγοντας δεν εξαρτάται από την ικανότητα του εξεταζόμενου, δηλαδή τόσο οι εξεταζόμενοι χαμηλού όσο και αυτοί υψηλού επιπέδου ικανότητας έχουν την ίδια πιθανότητα να απαντήσουν σωστά στο αντικείμενο μαντεύοντας τη σωστή απάντηση.
Σε κάποιες παραλλαγές του μοντέλου αυτού ο παράγοντας τύχης δεν είναι σταθερός σε όλα τα επίπεδα ικανότητας, αλλά λαμβάνεται υπόψη ότι ο παράγοντας τύχης παίζει σημαντικότερο ρόλο σε εξεταζόμενους με χαμηλό επίπεδο από ότι σε αυτούς με υψηλό επίπεδο ικανότητας. Θεωρητικά το πεδίο τιμών του παράγοντα c είναι (0,1) επειδή εκφράζει μια πιθανότητα, πρακτικά όμως οι τυπικές τιμές του κυμαίνονται στο διάστημα (0, 0.35).

Στο Σχήμα 4 φαίνεται η χαρακτηριστική καμπύλη ενός αντικειμένου του μοντέλου 3PL, με παραμέτρους a=1, b=0.5, c=0.2. Όπως φαίνεται και από το σχήμα το σημείο του άξονα ικανοτήτων για το οποίο ισχύει P(θ)=0,5 είναι το θ=0. Προφανώς ο ορισμός που είχε δοθεί για το b στα δύο προηγούμενα μοντέλα δεν ισχύει σε αυτό. Σε αυτό το μοντέλο ο παράγοντας δυσκολίας b αντιστοιχεί στο σημείο του άξονα ικανοτήτων για το οποίο ισχύει

$$P(\theta) = \frac{c + (1 - c)}{2} \quad (\Sigma. 5)$$

Η πιθανότητα αυτή είναι το μέσο μεταξύ των τιμών P(θ)=c και P(θ)=1. Αυτό που έχει αλλάξει στο μοντέλο αυτό είναι ότι ο παράγοντας τύχης c ορίζει το κάτω όριο στο οποίο πλησιάζει ασυμπτωτικά η καμπύλη. Η χαμηλότερη δυνατή πιθανότητα δηλαδή δεν είναι πλέον μηδέν αλλά c. Έτσι ο παράγοντας δυσκολίας ορίζεται από το σημείο του άξονα των ικανοτήτων που βρίσκεται στο μέσο του διαστήματος [c,1].

Ο παράγοντας διακριτικής ικανότητας a είναι και σε αυτό το μοντέλο ανάλογος της κλίσης της καμπύλης στο σημείο θ=b, σε αυτή την περίπτωση όμως η κλίση στο σημείο αυτό είναι $\frac{a(1-c)}{4}$. Αν και οι αλλαγές από το μοντέλο δύο παραμέτρων όσο αφορά τις παραμέτρους a και b φαίνονται πολύ μικρές, κατά την εκτίμηση των πειραματικών αποτελεσμάτων τα δύο μοντέλα διαφέρουν κατά πολύ.



Σχήμα 4 : Η χαρακτηριστική καμπύλη αντικειμένου για ένα αντικείμενο μοντέλου 3PL με παραμέτρους a=2 , b=1 , c=0,2

Για αντικείμενα του μοντέλου μικρής δυσκολίας και διακριτικής ικανότητας(όταν b<0 και a<1) η επίδραση του παράγοντα τύχης c μειώνεται στην περίπτωση που μελετάμε επίπεδα ικανότητας στο διάστημα θ∈[-3,3]. Αυτό συμβαίνει γιατί σε αυτές τις περιπτώσεις ο παράγοντας c δεν ορίζει το κάτω άκρο της πιθανότητας σωστής απάντησης, επειδή η κλίση της καμπύλης είναι μικρή. Αν μελετούσαμε μεγαλύτερο διάστημα του επιπέδου ικανότητας η καμπύλη θα πλησίαζε ασυμπτωτικά το c καθώς το θ πλησιάζει το -∞.

Όπως φαίνεται και από τον τύπο της P(θ) οι ορισμοί ενός μοντέλου 3PL με παράμετρο c=0 συμπίπτουν με αυτούς του μοντέλου 2PL.

Β.1.4 Μοντέλα με Πολυχότομα αντικείμενα.

Τα CAT που χρησιμοποιούν διχότομα αντικείμενα δεν χρησιμοποιούν τα συμπεράσματα μερικών ερευνών σχετικά με την ανθρώπινη γνώση όπως είναι οι έρευνες των Brown και Burton (1978) , Brown και VanLehn (1980) , Lane Stone και Hsu (1990) , Tatsuoka (1983).
Για παράδειγμα η ανάλυση του Tatsuoka για την αντίληψη των μαθητών στην επίλυση μαθηματικών προβλημάτων δείχνει ότι οι λανθασμένες απαντήσεις μπορούν να είναι πολλών ειδών. Παρ' όλα αυτά όταν χρησιμοποιούμε διχότομα αντικείμενα όλες οι λανθασμένες απαντήσεις βαθμολογούνται αδιακρίτως με μηδέν. Επιπλέον ο Nedelsky (1954) έδειξε χρησιμοποιώντας την κλασική θεωρία εξέτασης, όπως και οι Levine και Drasgow (1983) χρησιμοποιώντας την IRT, ότι η κατανομή των λανθασμένων απαντήσεων στις διάφορες λανθασμένες επιλογές όταν χρησιμοποιούνται ερωτήσεις με πολλαπλές εναλλακτικές απαντήσεις διαφέρουν σε διαφορετικά επίπεδα ικανότητας των εξεταζόμενων. Επιπλέον έδειξαν ότι οι λανθασμένες εναλλακτικές απαντήσεις μπορούν να προσφέρουν πληροφορία σχετικά με το βαθμό κατανόησης της ερώτησης από τον εξεταζόμενο. Με αυτό τον τρόπο οι λανθασμένες εναλλακτικές μπορούν να βοηθήσουν στην αύξηση της ακρίβειας της εκτίμησης της ικανότητας του εξεταζόμενου.

Χρησιμοποιώντας διχότομα αντικείμενα δεν λαμβάνουμε υπόψη τις λανθασμένες εναλλακτικές απαντήσεις. Όλες βαθμολογούνται με μηδέν και έτσι δεν αντλούμε καμία πληροφορία από αυτές σχετικά με την ικανότητα του εξεταζόμενου. Αντίθετα με τη χρήση πολυχότομων αντικειμένων μπορούμε να εκμεταλλευτούμε την πληροφορία που παρέχεται από τις λανθασμένες εναλλακτικές απαντήσεις. Σε αυτή την περίπτωση στο κάθε αντικείμενο η σωστή απάντηση βαθμολογείται με κάποιο βαθμό, ενώ η κάθε λανθασμένη με ένα μικρότερο βαθμό ή μηδέν. Υπάρχουν δύο κατηγορίες μοντέλων IRT που χρησιμοποιούν πολυχότομα αντικείμενα και χρησιμοποιούνται στα CATs, τα διαφορικά μοντέλα(difference models) και τα μοντέλα υποδιαίρεσης του συνόλου (divide-by-total models).

Β.1.4.1 Το μοντέλο βαθμιαίας απάντησης (graded response model)

Το βασικό διαφορικό μοντέλο είναι το μοντέλο βαθμιαίας απάντησης (graded response model). Τα υπόλοιπα μοντέλα αυτής της κατηγορίας (π.χ. Rating scale model) βασίζονται σε αυτό. Το Graded Response είναι μια επέκταση του 2PL διχότομου μοντέλου για πολύχοτομα αντικείμενα. Αντιπροσωπεύει μια σειρά μαθηματικών μοντέλων που αναπαριστούν διατεταγμένες πολύτομες κατηγορίες. Τέτοια μοντέλα είναι κατάλληλα π.χ. για μια έρευνα προσωπικότητας με εναλλακτικές απαντήσεις του τύπου «διαφωνώ πλήρως», «διαφωνώ», «συμφωνώ» και «συμφωνώ απόλυτα» ή η βαθμολογία μιας μερικά σωστής απάντησης ή η βαθμολογία με γράμματα π.χ. Α, Β , Γ στην εκτίμηση της επίδοσης των μαθητών. Για κάθε αντικείμενο ορίζονται κάποιες κατηγορίες(στάδια) οι οποίες βαθμολογούνται με ένα μέρος του συνολικού βαθμού του αντικειμένου. Με αυτό το μοντέλο μπορούμε δηλαδή να χρησιμοποιήσουμε ερωτήσεις που περιλαμβάνουν περισσότερα από ένα στάδια. Είναι λοιπόν καταλληλότερο σε ένα CAT όταν οι ερωτήσεις περιλαμβάνουν προβλήματα μαθηματικών, φυσικής κτλ. Έτσι για μια ερώτηση που βαθμολογείται με 5 μονάδες ο

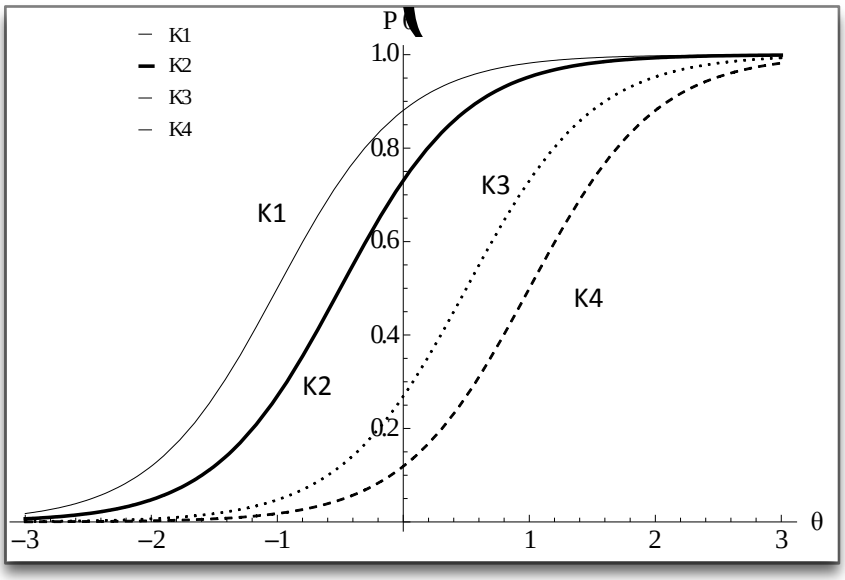
εξεταζόμενος μπορεί να κερδίζει μια μονάδα με την επιτυχή συμπλήρωση ενός βήματος της ακολουθίας που απαιτείται για την επίλυση του προβλήματος.
Για κάθε αντικείμενο εκτιμάται μια παράμετρος διάκρισης και τα όρια της κατηγορίας που ανήκει το αντικείμενο. Ο Samejima, ο οποίος ήταν αυτός που ανέπτυξε αυτό το μοντέλο το 1969, πρότεινε μια διαδικασία δύο βημάτων για να προσδιοριστεί η πιθανότητα κάποιος εξεταζόμενος να έχει κάποιο βαθμό σε ένα αντικείμενο i. Παραλλάσσοντας αυτή τη διαδικασία προκύπτουν τα περισσότερα από τα υπόλοιπα μοντέλα της κατηγορίας αυτής. Το Graded Response εκπίπτει στο 2PL μοντέλο αν ορίσουμε δύο κατηγορίες απαντήσεων.

Η πιθανότητα ότι η βαθμολογία σε μια κατηγορία κάποιου εξεταζόμενου επιπέδου ικανότητας θ θα είναι μεγαλύτερη ή ίση από ένα συγκεκριμένο βαθμό σε ένα αντικείμενο i δίνεται από τον τύπο

$$P_{x_i}{}^{*}(\theta) = \frac{e^{\alpha_{x_i}(\theta - b_{x_i})}}{1 + e^{\alpha_{x_i}(\theta - b_{x_i})}} \quad (\Sigma.6)$$

Όπου
b_{x_i} είναι το όριο της κατηγορίας που αντιστοιχεί σε μια συγκεκριμένη βαθμολογία x_i (με x_i=1,… m_i) όπου m_i είναι ο αριθμός των κατηγοριών του αντικειμένου i
a ο παράγοντας διακριτικής ικανότητας του αντικειμένου i
θ το επίπεδο ικανότητας

Στην ουσία χρησιμοποιώντας αυτό το μοντέλο αντιμετωπίζουμε το κάθε πολυχότομα βαθμολογημένο αντικείμενο σαν περισσότερα διατεταγμένα διχότομα βαθμολογημένα.
Η κάθε κατηγορία σε ένα αντικείμενο έχει τη δική της χαρακτηριστική καμπύλη. Στο Σχήμα 5 φαίνονται οι χαρακτηριστικές καμπύλες κατηγοριών για ένα αντικείμενο αυτού του μοντέλου με 4 κατηγορίες του οποίου η μέγιστη συνολική βαθμολογία είναι 4.



Σχήμα 5: οι χαρακτηριστικές καμπύλες κατηγοριών για ένα πούχότομο αντικείμενο με 4 κατηγορίες.

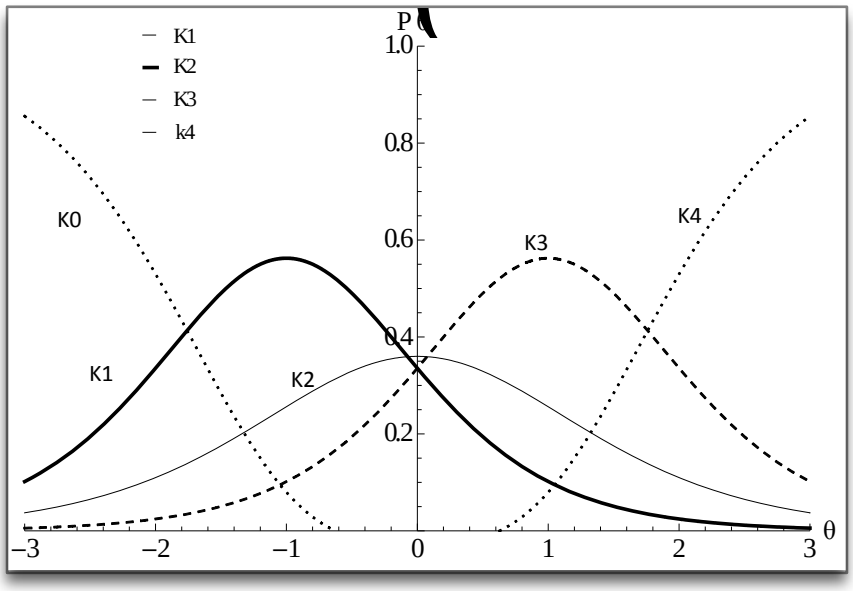
Ουσιαστικά στο παραπάνω παράδειγμα η καμπύλη K1 που αντιστοιχεί στην πρώτη κατηγορία εκφράζει την πιθανότητα ένας εξεταζόμενος ικανότητας θ να απαντήσει σωστά

στην πρώτη κατηγορία, η Κ2 στην πρώτη και τη δεύτερη, η Κ3 στις τρείς πρώτες και η Κ4 και στις τέσσερις. Δηλαδή, αν υποθέσουμε ότι οι κατηγορίες βαθμολογούνται διχότομα, η Κ1 δίνει την πιθανότητα ένας εξεταζόμενος να πάρει βαθμό για αυτό το αντικείμενο μεγαλύτερο ή ίσο με 1, η Κ2 μεγαλύτερο ίσο του 2 κ.ο.κ.

Για να υπολογίσουμε την πιθανότητα ένας εξεταζόμενος να απαντήσει σωστά σε μια κατηγορία και όχι στις επόμενες, την πιθανότητα δηλαδή που μας ενδιαφέρει, απαιτείται ένα επιπλέον βήμα. Σύμφωνα με τη θεωρία του Samejima η πιθανότητα αυτή δίνεται από τον τύπο

$$P_{x_i}(\theta) = P_{x_i}{}^{*}(\theta) - P_{x_i+1}{}^{*}(\theta) \quad (\Sigma.7)$$

Από τον τύπο αυτό παίρνουμε τις χαρακτηριστικές καμπύλες ενός αντικειμένου στο μοντέλο αυτό. Οι καμπύλες αυτές είναι αυτές που χρησιμοποιούνται στην πράξη. Για να πάρουμε την πιθανότητα σωστής απάντησης στην πρώτη κατηγορία και όχι στις επόμενες, η χαρακτηριστική της πρώτης κατηγορίας αφαιρείται από το 1. Όμοια για την πιθανότητα απάντησης στην τελευταία κατηγορία αφαιρούμε το μηδέν από τη χαρακτηριστική της κατηγορίας αυτής. Στο Σχήμα 6 φαίνονται οι λειτουργικές χαρακτηριστικές καμπύλες του αντικειμένου του Σχήματος 5.



Σχήμα 6 : Οι λειτουργικές χαρακτηριστικές καμπύλες ενός αντικειμένου με τέσσερις κατηγορίες

Ουσιαστικά οι πιθανότητες P_k του Σχήματος 6 εκφράζουν την πιθανότητα ένας εξεταζόμενος ικανότητας θ να φτάσει στο στάδιο κ, δηλαδή να απαντήσει σωστά μέχρι και το στάδιο κ.

Β.1.4.2 Το μοντέλο ονομαστικής απάντησης (nominal response model ή NR model)

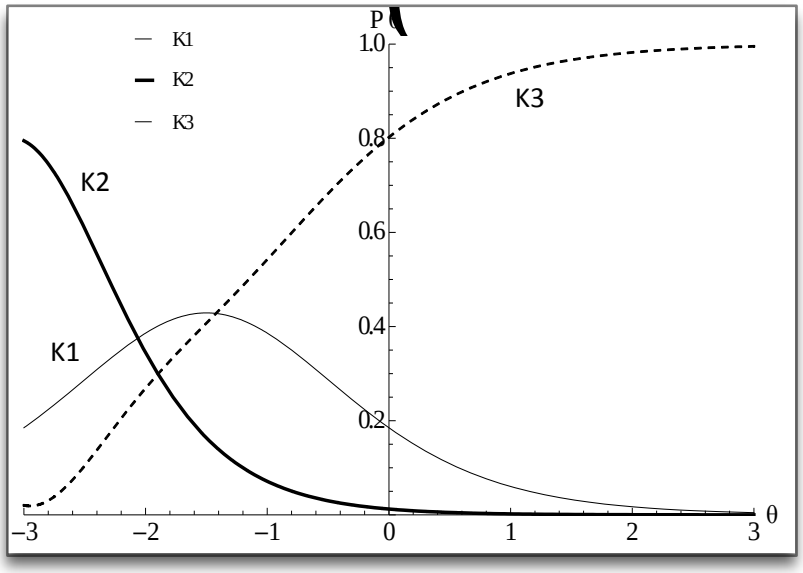
Το βασικό μοντέλο υποδιαίρεσης συνόλου είναι το μοντέλο ονομαστικής απάντησης (nominal response model). Όπως και στο graded response model οι εναλλακτικές απαντήσεις χωρίζονται σε κατηγορίες, και η κάθε κατηγορία βαθμολογείται με όλο το βαθμό που αντιστοιχεί στο αντικείμενο, ένα μέρος του όλου βαθμού ή μηδέν. Στο μοντέλο αυτό όμως οι κατηγορίες δεν είναι διατεταγμένες. Χρησιμοποιείται σε περιπτώσεις που είναι δύσκολο να διατάξουμε τις εναλλακτικές απαντήσεις ανάλογα με την ποσότητα της μερικής γνώσης που πρέπει να κατέχει ένα άτομο για να αναγνωρίσει ότι κάποια εναλλακτική απάντηση είναι λανθασμένη. Με τη χρήση του μοντέλου αυτού πετυχαίνουμε μεγαλύτερη ακρίβεια στις εκτιμήσεις μας σε κάποιες περιπτώσεις αφού χρησιμοποιείται η μερική γνώση στην περίπτωση μιας λανθασμένης απάντησης. Όπως και το graded response, και αυτό το μοντέλο εκπίπτει στο 2PL διχότομο μοντέλο όταν σε κάθε αντικείμενο ορίσουμε δύο κατηγορίες απαντήσεων (σωστό με βαθμολογία 1 και λάθος με βαθμολογία 0). Τα υπόλοιπα μοντέλα divide-by-total προκύπτουν από το nominal response model. Μοντέλα που ανήκουν στην κατηγορία αυτή είναι τα : Generalized Partial Credit, Partial Credit, Successive Intervals, Rating Scale και τα παράγωγά τους.

Στο μοντέλο NR η πιθανότητα η απάντηση ενός εξεταζόμενου ικανότητας θ να ανήκει στην κατηγορία j για το αντικείμενο i δίνεται από τον τύπο

$$P_{ij}(\theta) = \frac{e^{c_{ij}+a_{ij}\theta}}{\sum_{h=i}^{m_i} e^{c_{ij}+a_{ij}\theta}} \quad (\Sigma.8)$$

Όπου
a είναι ο παράγοντας διακριτικής ικανότητας
c_{ij} είναι ο παράγοντας αποκοπής της μη γραμμικής συνάρτησης που σχετίζεται με την κατηγορία j του αντικειμένου i
m_i είναι ο αριθμός των κατηγοριών του αντικειμένου i

Η πιθανότητα να ανήκει μια απάντηση σε μια κατηγορία συναρτήσει της ικανότητας φαίνεται στη χαρακτηριστική καμπύλη κατηγοριών του αντικειμένου. Στο σχήμα 7 φαίνονται οι χαρακτηριστικές καμπύλες κατηγοριών για ένα αντικείμενο του μοντέλου αυτού που έχει 3 κατηγορίες απαντήσεων. Η καμπύλη Κ1 απεικονίζει την πιθανότητα η απάντηση να ανήκει στην πρώτη κατηγορία, η Κ2 στη δεύτερη και η Κ3 στην τρίτη.



Σχήμα 7 : Οι χαρακτηριστικές καμπύλες κατηγοριών ενός αντικειμένου με τρεις κατηγορίες του NR μοντέλου.

Β.1.5 Πολυδιάστατα μοντέλα IRT

Όλα τα μοντέλα που έχουν αναλυθεί στις προηγούμενες παραγράφους, τα οποία είναι αυτά που χρησιμοποιούνται τις περισσότερες φορές όταν εφαρμόζεται η IRT θεωρούν την καθεμιά από τις ικανότητες του εξεταζόμενου ανεξάρτητη από τις άλλες. Αυτό όμως δεν είναι πάντοτε απόλυτα σωστό. Η IRT αποτελεί ως επί τω πλείστω μια μονοδιάστατη προσέγγιση στην εκτίμηση των γνωρισμάτων των εξεταζόμενων. Για το λόγο αυτό η θεωρία αυτή χρησιμοποιούνταν κυρίως στον τομέα της εκπαίδευσης.

Τα τελευταία χρόνια όμως, έχουν γίνει σημαντικές έρευνες που επιτρέπουν την εφαρμογή της θεωρίας σε τομείς που απαιτούν πολυδιάστατες προσεγγίσεις για την εκτίμηση των γνωρισμάτων των ατόμων, όπως η υγεία και η ψυχολογία. Εξετάσεις που βασίζονται σε τέτοιες έρευνες, είναι σχεδιασμένες ώστε να δίνουν εκτενείς πληροφορίες σχετικά με πολλά επίπεδα γνώσης, κατάστασης της υγείας, συμπεριφοράς ή άλλων χαρακτηριστικών της προσωπικότητας ενός ατόμου. Για παράδειγμα οι εξετάσεις GRE αποτελούνται από πολλά επιμέρους διαγωνίσματα με σκοπό την εκτίμηση των ικανοτήτων των εξεταζόμενων σε διάφορους τομείς, χωρίς απαραίτητα αυτοί οι τομείς να είναι ασυσχέτιστοι μεταξύ τους. Η ικανότητα δηλαδή ενός εξεταζόμενου σε ένα τομέα θα μπορούσε να τον βοηθήσει να απαντήσει σωστά σε κάποια αντικείμενα ενός διαγωνίσματος για την εκτίμηση της ικανότητάς του σε έναν άλλο τομέα. Ένας εξεταζόμενος που έχει υψηλό επίπεδο ικανότητας για παράδειγμα στα μαθηματικά είναι πολύ πιθανό να έχει υψηλό επίπεδο και στη φυσική καθώς οι δυο επιστήμες έχουν μεγάλη συσχέτιση. Είναι λοιπόν φυσικό να συμπεράνει κάποιος ότι και η ικανότητα στον ένα τομέα έχει μεγάλη συσχέτιση με την ικανότητα στον άλλο. Παρόμοια έχουν εφαρμοστεί εξετάσεις στο χώρο της υγείας με σκοπό την εκτίμηση της κατάστασης της υγείας ενός ατόμου, που αποτελούνται από πολλές διαφορετικές επιμέρους εξετάσεις που αφορούν διαφορετικούς τομείς της υγείας.

Όπως έχει αναφερθεί για την ανάπτυξη τέτοιων διαγωνισμάτων, που περιλαμβάνουν πολλούς συσχετισμένους τομείς , τα κλασικά μοντέλα της IRT δεν αρκούν. Θα πρέπει να

αναπτυχθούν να αναλυθούν και να εφαρμοστούν μοντέλα που θα λαμβάνουν υπ' όψη τη συσχέτιση των επιμέρους ικανοτήτων και την αλληλεπίδρασή τους.

Ήδη από τη δεκαετία του 80 έγιναν έρευνες για την εύρεση μοντέλων IRT για την πολυδιάστατη ανάλυση των δειγμάτων στην εκτίμηση της ικανότητας των ατόμων. Τέτοια μοντέλα είναι το πολυδιάστατο μοντέλο μιας παραμέτρου Reckase (1972) και το γραμμικό πολυδιάστατο μοντέλο δύο παραμέτρων McKinley & Reckase (1982). Αυτές και άλλες παρόμοιες μελέτες προσπαθούν να λάβουν υπ' όψη τη συσχέτιση που μπορούν να έχουν οι διάφορες ικανότητες ενός ατόμου, κατά την εκτίμηση των ικανοτήτων αυτών. Σύμφωνα με τη μελέτη του Segall (1996) χρησιμοποιώντας κάποιο πολυδιάστατο μοντέλο για την εκτίμηση ενός συνόλου ικανοτήτων μπορεί να επιτευχθεί μεγαλύτερη ακρίβεια εκτίμησης από ότι αν χρησιμοποιούνταν κάποιο μονοδιάστατο για κάθε μια από τις ιδιότητες ξεχωριστά. Τα τελευταία χρόνια έχουν προταθεί πολλά πολυδιάστατα μοντέλα. Όλα όμως περιορίζονται στη χρήση διχότομων αντικειμένων. Η χρήση πολυχότομων αντικειμένων σε τέτοια μοντέλα βρίσκεται ακόμη σε πειραματικό στάδιο.

Β.2 Η χαρακτηριστική καμπύλη του διαγωνίσματος (Test Characteristic Curve)

"Όταν βαθμολογείται ένα διαγώνισμα οι απαντήσεις του εξεταζόμενου στις ερωτήσεις του διαγωνίσματος βαθμολογούνται ξεχωριστά. Όταν τα αντικείμενα βαθμολογούνται διχότομα κάθε σωστή απάντηση παίρνει 1 και κάθε λανθασμένη 0. Όταν έχουμε πολυχότομη βαθμολόγηση η σωστή απάντηση παίρνει το μέγιστο βαθμό που αντιστοιχεί στην ερώτηση και καθεμιά από τις λανθασμένες απαντήσεις ένα μέρος του βαθμού αυτού. Η **ακατέργαστη βαθμολογία** του εξεταζόμενου στο διαγώνισμα αυτό **(raw test score)** μπορεί να υπολογιστεί αθροίζοντας τις βαθμολογίες που έχει λάβει στην κάθε ερώτηση. Η ακατέργαστη βαθμολογία είναι πάντα ένας ακέραιος αριθμός. Όταν τα αντικείμενα βαθμολογούνται διχότομα η βαθμολογία αυτή θα είναι από 0 ως Ν, όπου Ν ο αριθμός των ερωτήσεων του διαγωνίσματος. Αν οι εξεταζόμενοι καλούνταν να απαντήσουν πάλι στο διαγώνισμα, και αν υποθέσουμε πως δε θυμούνται τις απαντήσεις που είχαν δώσει την προηγούμενη φορά, θα παίρναμε διαφορετικές ακατέργαστες βαθμολογίες για τον κάθε εξεταζόμενο από την πρώτη φορά που είχε συμπληρώσει το διαγώνισμα. Υποθετικά ένας εξεταζόμενος θα μπορούσε να συμπληρώσει το διαγώνισμα πολλές φορές. Σε αυτή την περίπτωση οι ακατέργαστες πληροφορίες που θα έπαιρνε δεν θα ήταν ίδιες. Είναι λογικό κάποιος να περιμένει αυτές οι βαθμολογίες να συγκλίνουν γύρω από μία μέση τιμή. Στη θεωρία μετρήσεων αυτή η τιμή ονομάζεται **πραγματική βαθμολογία (true score)** και ο ορισμός της εξαρτάται από τη θεωρία μέτρησης που χρησιμοποιούμε.

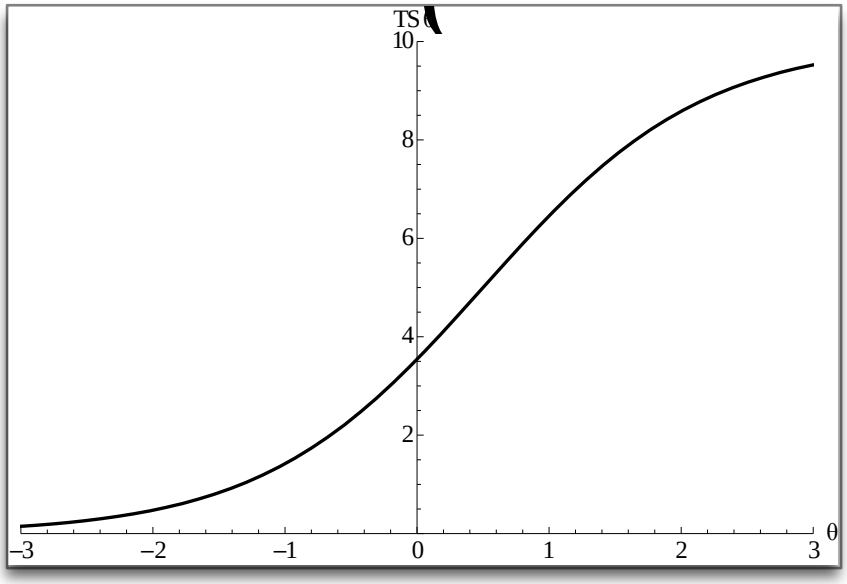
Στην IRT χρησιμοποιείται ένας ορισμός για την πραγματική βαθμολογία που δόθηκε από τον D.N Lawley. Ο τύπος για τον υπολογισμό της πραγματικής βαθμολογίας σύμφωνα με αυτή τη θεωρία είναι

$$TS_j = \sum_{i=1}^{N} P_i(\theta_j) \quad (\Sigma. 9)$$

Όπου TS_j είναι η πραγματική βαθμολογία για εξεταζόμενους επιπέδου ικανότητας θ_j
Το i είναι ο δείκτης των αντικειμένων
Το $P_i(\theta_j)$ είναι η πιθανότητα σωστής απάντησης για εξεταζόμενους επιπέδου ικανότητας θ_j και εξαρτάται από το μοντέλο χαρακτηριστικής καμπύλης αντικειμένου που χρησιμοποιούμε.

Αν για παράδειγμα έχουμε εκτιμήσει ότι η ικανότητα ενός εξεταζόμενου είναι $\hat{\theta}$=1,5 από ένα διαγώνισμα τεσσάρων ερωτήσεων στο οποίο έχει απαντήσει, μπορούμε από τις χαρακτηριστικές καμπύλες του κάθε αντικειμένου να υπολογίσουμε τις πιθανότητες να απαντήσει σωστά κάποιος εξεταζόμενος ικανότητας $\hat{\theta}$=1,5 στην κάθε ερώτηση. Αν αυτές ήταν $P_1(1,5)=0,7$, $P_2(1,5)=0,3$, $P_3(1,5)=0,6$ και $P_4(1,5)=0,9$ τότε η πραγματική βαθμολογία του εξεταζόμενου θα ήταν TS = 0,7 + 0,3 + 0,6 + 0,9 = 2,5. Αυτή θα ήταν η πραγματική βαθμολογία κάθε εξεταζόμενου του οποίου έχουμε εκτιμήσει ότι η ικανότητα είναι $\hat{\theta}$=1,5.
Αν και κανένας εξεταζόμενος δεν πρόκειται να έχει ακατέργαστη βαθμολογία 2,5 (υποθέτοντας ότι οι ερωτήσεις βαθμολογούνται διχότομα) αυτή είναι θεωρητικά η μέση τιμή όλων των ακατέργαστων βαθμολογιών που θα έπαιρναν εξεταζόμενοι ικανότητας θ=1,5 σε αυτό το διαγώνισμα, αν το είχαν συμπληρώσει πολλές φορές.

Η χαρακτηριστική καμπύλη του διαγωνίσματος είναι η γραφική αναπαράσταση της πραγματικής βαθμολογίας συναρτήσει της ικανότητας. Η πραγματική βαθμολογία έχει πεδίο τιμών [0,N], όπου N ο αριθμός των ερωτήσεων στο διαγώνισμα. Στο Σχήμα 5 φαίνεται μια τυπική χαρακτηριστική καμπύλη ενός διαγωνίσματος δέκα ερωτήσεων.



Σχήμα 8 : Μια τυπική χαρακτηριστική καμπύλη διαγωνίσματος ενός διαγωνίσματος 10 ερωτήσεων.

Από αυτή την καμπύλη αν έχουμε το επίπεδο ικανότητας ενός εξεταζόμενου μπορούμε να υπολογίσουμε την πραγματική βαθμολογία του και το αντίστροφο. Αν για παράδειγμα

έχουμε εκτιμήσει την ικανότητα κάποιου εξεταζόμενου $\hat{\theta}$=0 από το διαγώνισμα του Σχήματος 8 τότε η πραγματική του βαθμολογία θα είναι 3,5 στα 10.

Αν χρησιμοποιήσουμε το μοντέλο μιας παραμέτρου ή το μοντέλο δύο παραμέτρων για ένα διαγώνισμα του οποίου οι ερωτήσεις βαθμολογούνται διχότομα, τότε η χαρακτηριστική καμπύλη του διαγωνίσματος πλησιάζει ασυμπτωτικά το μηδέν καθώς η ικανότητα πλησιάζει το -∞. Όταν η ικανότητα πλησιάζει το +∞ η χαρακτηριστική πλησιάζει το N, όπου N ο αριθμός των ερωτήσεων του διαγωνίσματος. Αν χρησιμοποιήσουμε το μοντέλο των τριών παραμέτρων για N διχότομα αντικείμενα σε ένα διαγώνισμα το κάτω άκρο της χαρακτηριστικής του διαγωνίσματος πλησιάζει το άθροισμα των παραμέτρων τύχης των αντικειμένων και όχι το μηδέν. Αυτό συμβαίνει γιατί, σύμφωνα με αυτό το μοντέλο, οι εξεταζόμενοι με χαμηλό επίπεδο ικανότητας μπορούν να απαντήσουν κάποιες ερωτήσεις σωστά μαντεύοντας τις σωστές απαντήσεις. Το πάνω άκρο της καμπύλης και σε αυτή την περίπτωση πλησιάζει ασυμπτωτικά τον αριθμό των ερωτήσεων του διαγωνίσματος.
Ο κύριος ρόλος της χαρακτηριστικής καμπύλης διαγωνίσματος στην IRT είναι να μας δώσει ένα τρόπο να μετατρέπουμε τις τιμές της ικανότητας σε πραγματική βαθμολογία. Αυτό είναι χρήσιμο για άτομα που δεν έχουν γνώση της IRT. Μετατρέποντας την ικανότητα σε πραγματική βαθμολογία, δίνουμε στον ενδιαφερόμενο τη βαθμολογία με έναν αριθμό σε σχέση με το συνολικό αριθμό των ερωτήσεων του διαγωνίσματος και έτσι η βαθμολογία γίνεται περισσότερο κατανοητή καθώς αυτή η αναπαράσταση είναι πιο οικεία στους ανθρώπους. Η χαρακτηριστική καμπύλη διαγωνίσματος επίσης παίζει σημαντικό ρόλο στις διαδικασίες σύγκρισης διαφορετικών διαγωνισμάτων.

Η γενική μορφή της χαρακτηριστικής καμπύλης διαγωνίσματος είναι αυτή μιας μονότονα αύξουσας συνάρτησης. Σε πολλές περιπτώσεις έχει σχήμα S, παρόμοιο με αυτό της χαρακτηριστικής καμπύλης αντικειμένου. Σε άλλες περιπτώσεις είναι αρχικά αύξουσα, έπειτα παραμένει περίπου σταθερή σε μια μικρή περιοχή και τέλος γίνεται και πάλι αύξουσα. Σε όλες τις περιπτώσεις όμως τελικά πλησιάζει ασυμπτωτικά τον αριθμό των ερωτήσεων του διαγωνίσματος στο άνω άκρο της. Η μορφή της εξαρτάται από πολλούς παράγοντες όπως είναι ο αριθμός των ερωτήσεων στο διαγώνισμα, το μοντέλο IRT που χρησιμοποιείται καθώς το κάθε μοντέλο έχει διαφορετική χαρακτηριστική καμπύλη αντικειμένων, και οι τιμές των παραμέτρων των αντικειμένων. Ο μόνος τρόπος να σχηματίσει κάποιος τη χαρακτηριστική καμπύλη ενός διαγωνίσματος είναι να υπολογίσει την πιθανότητα σωστής απάντησης για κάθε επίπεδο ικανότητας και για όλα τα αντικείμενα του διαγωνίσματος χρησιμοποιώντας τις χαρακτηριστικές καμπύλες αντικειμένων του μοντέλου IRT που έχει χρησιμοποιηθεί. Η μορφή της χαρακτηριστικής του διαγωνίσματος δεν εξαρτάται από την κατανομή της συχνότητας εμφάνισης των ικανοτήτων. Όπως και η χαρακτηριστική αντικειμένου, έτσι και η χαρακτηριστική διαγωνίσματος απεικονίζουν συναρτήσεις μεταξύ δύο κλιμάκων και δεν εξαρτώνται από την κατανομή των βαθμολογιών στις κλίμακες αυτές.

Η τιμή της ικανότητας για την οποία γίνεται $TS = \frac{N}{2}$ εκφράζει τη συνολική δυσκολία ενός διαγωνίσματος N ερωτήσεων.

Β.3 Εκτίμηση της ικανότητας του εξεταζόμενου.

Ο κύριος σκοπός ενός CAT όπως και της IRT είναι να τοποθετήσει τον εξεταζόμενο στην κλίμακα των ικανοτήτων. Πετυχαίνοντας αυτό μπορούμε εκτός από την εκτίμηση της ικανότητας του εξεταζόμενου και να κάνουμε συγκρίσεις μεταξύ των εξεταζόμενων εφόσον τις χρειαζόμαστε π.χ. για να υπολογίσουμε τη βαθμολογία τους σε ένα μάθημα ή να βρούμε τους καλύτερους οι οποίοι θα πάρουν κάποιο βραβείο.

Για τον υπολογισμό της ικανότητας του εξεταζόμενου χρησιμοποιείται ένα CAT που αποτελείται από Ν αντικείμενα, καθένα από τα οποία μετρά ένα μέρος της ικανότητας αυτής. Για την πραγματοποίηση ενός τέτοιου διαγωνίσματος πρέπει αρχικά να έχουμε επιλέξει ένα μοντέλο IRT ανάλογα με τις απαιτήσεις της εξέτασης και τη μορφή των ερωτήσεων, να έχουμε δημιουργήσει μια βάση με τις ερωτήσεις του διαγωνίσματος και να έχουμε εκτιμήσει τις απαραίτητες παραμέτρους για το κάθε αντικείμενο στη βάση μας. Αν υποθέσουμε ότι όλα αυτά έχουν ήδη γίνει, οι παράμετροι για το κάθε αντικείμενο είναι γνωστές. Συνεπώς η κλίμακα μέτρησης της ικανότητας στις εκτιμήσεις της ικανότητας θα είναι η ίδια με αυτή που χρησιμοποιείται από τις παραμέτρους, δηλαδή το διάστημα [-3,3]. Οι βαθμολογίες που παίρνει ο εξεταζόμενος σε κάθε ένα από τα Ν αντικείμενα του διαγωνίσματος εισάγονται σε ένα άνυσμα που ονομάζεται άνυσμα απαντήσεων εξεταζόμενου. Αυτό που θέλουμε να πετύχουμε είναι να εκτιμήσουμε την άγνωστη ικανότητα του εξεταζόμενου χρησιμοποιώντας το άνυσμα αυτό και τις παραμέτρους των αντικειμένων στα οποία έχει απαντήσει ο εξεταζόμενος. Υπάρχουν πολλές μέθοδοι για τον υπολογισμό της εκτίμησης αυτής. Οι κυριότερες είναι η μέθοδος μέγιστης πιθανοφάνειας (maximum likelihood estimation ή mle), η μέθοδος οριακής πιθανοφάνειας (marginal maximum likelihood estimation) και μέθοδοι εκτίμησης που χρησιμοποιούν μοντέλα Bayess. Αυτή που χρησιμοποιείται συχνότερα είναι η μέθοδος μέγιστης πιθανοφάνειας. Η μέθοδος αυτή χρησιμοποιήθηκε και στην εργασία αυτή.

Β.3.1 Η μέθοδος μέγιστης πιθανοφάνειας (maximum likelihood estimation).

Η μέθοδος μέγιστης πιθανοφάνειας είναι μια επαναληπτική μέθοδος που χρησιμοποιείται στη στατιστική για την εκτίμηση των παραμέτρων μιας συνάρτησης τυχαίας μεταβλητής. Σε αρχικό στάδιο ξεκινά με κάποια τιμή στην οποία έχουμε αρχικοποιήσει εμείς την ικανότητα του εξεταζόμενου, η οποία είναι η παράμετρος της συνάρτησης που θέλουμε να εκτιμήσουμε. Μετά από κάθε απάντηση, υπολογίζεται μια νέα τιμή για την εκτιμούμενη ικανότητα του εξεταζόμενου χρησιμοποιώντας την ήδη υπάρχουσα εκτιμούμενη ικανότητα, τις παραμέτρους των αντικειμένων στα οποία έχει απαντήσει, και το άνυσμα των απαντήσεών του. Ουσιαστικά υπολογίζονται οι πιθανότητες σωστής απάντησης για τα αντικείμενα που έχει απαντήσει ο εξεταζόμενος. Έπειτα υπολογίζεται μια διόρθωση της εκτίμησης της ικανότητας, ώστε να ταιριάζουν όσο το δυνατό καλύτερα οι πιθανότητες που έχουν υπολογιστεί με το άνυσμα απαντήσεων του εξεταζόμενου. Η διαδικασία επαναλαμβάνεται ώσπου η διόρθωση αυτή να είναι πολύ μικρή, οπότε οι αλλαγές στην εκτιμούμενη ικανότητα στις επόμενες επαναλήψεις θα είναι αμελητέες. Το αποτέλεσμα

είναι η ζητούμενη εκτίμηση της ικανότητας. Η διαδικασία αυτή γίνεται σε ένα CAT για κάθε εξεταζόμενο ξεχωριστά και κάθε φορά που απαντά σε μια ερώτηση.

Γνωρίζοντας τις παραμέτρους των αντικειμένων (a,b,c) και την ικανότητα ενός εξεταζόμενου j μπορούμε να υπολογίζουμε την υπό συνθήκη πιθανότητα του ανύσματος απαντήσεων X_i. Αυτή για κάθε παράγοντα του αντικειμένου i του ανύσματος δίνεται από τον τύπο :

$$P\left(X_{ij}=1\middle|\theta_j,\Omega_i\right)=P_{ij}Q_{ij}=P_{ij}\left(1-P_{ij}\right)=\frac{e^{-a(\theta-b)}}{1+e^{-a(\theta-b)}}$$

Η μέθοδος μέγιστης πιθανοφάνειας μας δίνει τις τιμές των παραμέτρων οι οποίες κάνουν το άνυσμα των απαντήσεων πιο πιθανό.

Η πιθανοφάνεια L στην ουσία εκφράζει την υπό συνθήκη πιθανότητα, όταν είναι γνωστές οι τιμές της τυχαίας μεταβλητής (στην περίπτωσή μας το άνυσμα των απαντήσεων) και οι παράμετροι μεταβάλλονται. Δηλαδή P(X|Θ,α,β)=L(Θ,α,β|X)

Η δεσμευμένη πιθανότητα L(Θ,α,β|X) γράφεται :

$$L(\Theta,\Omega|X)=\prod_{i=1}^{N}\prod_{j=1}^{M}P\left(X_{ij}=x_{ij}\middle|\theta_j,\alpha_i,b_i\right)=\prod_{i=1}^{N}\prod_{j=1}^{M}(P_{ij})^{x_{ij}}\,(1-P_{ij})^{1-x_{ij}}$$

Όπου x_{ij} είναι η βαθμολογία (0 ή 1) του εξεταζόμενου i στο αντικείμενο j. P_{ij} η πιθανότητα σωστής απάντησης , Θ_i η ικανότητα του εξεταζόμενου i και Ω_j το σύνολο των παραμέτρων του αντικειμένου j (a,b,c για το μοντέλο 3PL)

Εκείνο που ενδιαφέρει είναι η τιμή της παραμέτρου θ που μεγιστοποιεί την πιθανοφάνεια. Για το σκοπό αυτό είναι αριθμητικά πιο εύκολο να υπολογιστεί η τιμή της παραμέτρου που μεγιστοποιεί το φυσικό λογάριθμο της πιθανοφάνειας :

$$\ln\left[L\left(\boldsymbol{\Theta},\boldsymbol{\Omega}\mid\mathbf{X}\right)\right]=\sum_{i=1}^{N}\sum_{j=1}^{J}x_{ij}\ln(P_{ij})+(1-x_{ij})\ln(1-P_{ij}) \qquad (\Sigma.10)$$

Αυτό επειδή ο φυσικός λογάριθμος της παράστασης παίρνει τη μέγιστη τιμή του για την ίδια τιμή της παραμέτρου θ για την οποία παίρνει τη μέγιστη τιμή της και η παράσταση. Η πολυπλοκότητα των απαιτούμενων υπολογισμών όμως είναι πολύ μικρότερη.

Η πιθανοφάνεια είναι πάντα μεγαλύτερη ή ίση με το μηδέν και έχει γραφική παράσταση όμοια με την κατανομή Gauss. Έχει ένα μοναδικό μέγιστο όπου μηδενίζεται η πρώτη παράγωγός της. Για τον υπολογισμό του μεγίστου αρκεί να βρεθεί αυτό το σημείο μηδενισμού. Η πρώτη παράγωγος της είναι:

$$\frac{\partial\ln\left(L(\Theta,\Omega|X)\right)}{\partial\theta}=\sum_{i=1}^{N}\sum_{j=1}^{M}a_i(x_{ij}-P_{ij}) \qquad (\Sigma.11)$$

Για τον υπολογισμό του σημείου μηδενισμού της χρησιμοποιείται η μέθοδος Newton-Raphson. Σύμφωνα με τη μέθοδο αυτή για τον υπολογισμό της ρίζας μιας συνάρτησης f(x) ακολουθείται η παρακάτω διαδικασία

- επιλέγεται μια αρχική τιμή για τη μεταβλητή $\chi=\chi_0$
- υπολογίζεται ένας διορθωτικός παράγοντας $dx=\frac{f(x)}{\frac{\partial f(x)}{\partial x}}=\frac{f(x)}{f'(x)}$

- η επόμενη υποψήφια τιμή του χ είναι $x_{t+1} = x_t - dx$
- η διαδικασία επαναλαμβάνεται μέχρι η dx να πάρει τιμή μικρότερη από ένα κατώφλι που ορίζεται σε πολύ μικρή τιμή, δηλαδή να έχουμε πολύ μικρές μεταβολές στην τιμή του χ.

Στην περίπτωσή μας $f(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{M} a_i(x_{ij} - P_{ij})$ και $f'(\theta) = -\sum_{i=1}^{N}\sum_{j=1}^{M}({a_i}^2 P_{ij}Q_{ij})$

Οι παραπάνω τύποι ισχύουν για την εκτίμηση των ικανοτήτων Μ εξεταζομένων. Για την εκτίμηση ενός μόνο εξεταζόμενου παραλείπονται από την f και την f' οι όροι $\sum_{j=1}^{M}$.

Οπότε ο μαθηματικός τύπος που δίνει την εκτίμηση της ικανότητας ενός εξεταζόμενου, αφού αυτός έχει απαντήσει την ερώτηση Ν όταν χρησιμοποιούνται διχότομα βαθμολογημένα αντικείμενα είναι:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{\sum_{i=1}^{N} \alpha_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} {a_i}^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)} \quad (\Sigma.12)$$

Όπου
$\hat{\theta}_t$ είναι η εκτιμούμενη ικανότητα του εξεταζόμενου στην επανάληψη t της μεθόδου.
a_i είναι ο παράγοντας διακριτικής ικανότητας του αντικειμένου i , i=1,2,…N
u_i είναι ο βαθμός του εξεταζόμενου στο αντικείμενο i . Στην περίπτωση που τα αντικείμενα είναι διχότομα, θα είναι u_i=1 αν ο εξεταζόμενος έχει απαντήσει σωστά στο αντικείμενο ή u_i=0 αν έχει απαντήσει λάθος.
$P_i(\hat{\theta}_t)$ είναι η πιθανότητα σωστής απάντησης σύμφωνα με τη χαρακτηριστική καμπύλη του μοντέλου που χρησιμοποιείται, για επίπεδο ικανότητας $\hat{\theta}_s$
$Q_i(\hat{\theta}_t)$ είναι η πιθανότητα λανθασμένης απάντησης . Στην περίπτωση των διχότομων αντικειμένων ισχύει $Q_i(\hat{\theta}_t) = 1 - P_i(\hat{\theta}_t)$

Αρχικά η $\hat{\theta}$ έχει μια αρχική τιμή η οποία επιλέγεται αυθαίρετα π.χ. το 0. Η πιθανότητα σωστής απάντησης για καθένα από τα Ν αντικείμενα που έχουν απαντηθεί υπολογίζεται χρησιμοποιώντας την τιμή αυτή, τις παραμέτρους των αντικειμένων που είναι γνωστές και τη χαρακτηριστική καμπύλη του κάθε αντικειμένου. Έπειτα υπολογίζεται ο δεύτερος όρος του αθροίσματος. Αυτός είναι ο διορθωτικός παράγοντας $d\hat{\theta}$. Αθροίζοντας το διορθωτικό παράγοντα με την εκτιμούμενη ικανότητα παίρνουμε μια νέα καλύτερη εκτίμηση της ικανότητας. Η νέα εκτίμηση χρησιμοποιείται στην επόμενη επανάληψη. Ο όρος $u_i - P_i(\hat{\theta}_t)$ στον αριθμητή του διορθωτικού παράγοντα είναι η διαφορά της βαθμολογίας που πήρε ο εξεταζόμενος σε αυτό το αντικείμενο από την πιθανότητα σωστής απάντησης στο επίπεδο ικανότητας έχει εκτιμηθεί ότι βρίσκεται αυτός. Συνεπώς όσο πλησιάζει η εκτιμούμενη ικανότητα στην πραγματική ικανότητα του εξεταζόμενου το άθροισμα των διαφορών $u_i - P_i(\hat{\theta}_t)$ θα γίνεται μικρότερο, δηλαδή ο αριθμητής θα γίνεται μικρότερος όπως και ο διορθωτικός παράγοντας. Ο σκοπός μας είναι να βρούμε την εκτίμηση της ικανότητας που δίνει τιμές $P_i(\hat{\theta}_t)$ οι οποίες να ελαχιστοποιούν το άθροισμα του αριθμητή. Όταν συμβεί αυτό η τιμή του διορθωτικού παράγοντα θα είναι τόσο μικρή που οι μεταβολές της

εκτίμησης της ικανότητας θα είναι αμελητέες. Αυτή θα είναι η νέα εκτίμηση της ικανότητας του εξεταζόμενου μετά την απάντηση μιας ερώτησης. Η συνάρτηση Σ.12 μπορεί να χρησιμοποιηθεί για την εκτίμηση της ικανότητας στα μοντέλα 1PL και 2PL . Για το μοντέλο 3PL χρησιμοποιείται μια τροποποίησή της.

Μια ένδειξη για την ακρίβεια της εκτίμησης που χρησιμοποιείται συχνά είναι το τυπικό σφάλμα (standard error). Υποθετικά ένας εξεταζόμενος θα μπορούσε να συμπληρώσει ένα διαγώνισμα πολλές φορές, χωρίς να θυμάται τις απαντήσεις που είχε δώσει τις προηγούμενες φορές και χωρίς να μελετήσει μεταξύ των διαγωνισμάτων ώστε να αλλάξει το επίπεδο ικανότητάς του. Μετά από κάθε εξέταση θα υπήρχε και μια εκτίμηση $\hat{\theta}$ της ικανότητας του. Το τυπικό σφάλμα είναι ένα μέτρο του πόσο μεταβάλλονται οι εκτιμούμενες τιμές γύρω από την πραγματική ικανότητα. Ο τύπος που δίνει το τυπικό λάθος στη μέθοδο αυτή για τα μοντέλα μίας και δύο παραμέτρων είναι

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^{N} a^2{}_i P(\hat{\theta}) Q(\hat{\theta})}} \qquad (\Sigma.13)$$

Ενώ για το μοντέλο τριών παραμέτρων είναι

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^{N} a^2{}_i \frac{Q(\hat{\theta})}{P(\hat{\theta})} \left[\frac{P(\hat{\theta}) - c}{1 - c}\right]^2}} \qquad (\Sigma.14)$$

Προφανώς όσο περισσότερα αντικείμενα χρησιμοποιηθούν σε ένα διαγώνισμα τόσο μικρότερο θα είναι το τυπικό σφάλμα, και συνεπώς η εκτίμηση πιο ακριβής.

Χρησιμοποιώντας τη μέθοδο της μέγιστης πιθανοφάνειας λαμβάνονται πολύ ακριβείς εκτιμήσεις. Η μέθοδος όμως έχει και μερικά γνωστά προβλήματα.

Υπάρχουν δύο περιπτώσεις που η μέθοδος αυτή δεν μπορεί να εκτιμήσει την ικανότητα του εξεταζόμενου. Όταν έχει απαντήσει σε όλες τις ερωτήσεις ενός διαγωνίσματος σωστά ή σε όλες λάθος. Στην πρώτη περίπτωση η εκτιμούμενη ικανότητα είναι $+\infty$ και στη δεύτερη $-\infty$.
Συνήθως τα προγράμματα έχουν σχεδιαστεί ώστε στην τελική εκτίμηση της ικανότητας, όταν εντοπίσουν τέτοιες περιπτώσεις να μη συνεχίζουν στην ανάλυση των αποτελεσμάτων, ή να δίνουν τη μέγιστη βαθμολογία σημειώνοντας με κάποιο τρόπο ότι η εκτίμηση είναι θεωρητικά άπειρο. Πρακτικά στις περιπτώσεις που ο εξεταζόμενος έχει απαντήσει σωστά σε όλες ή σε καμία ερώτηση σημαίνει ότι το συγκεκριμένο τεστ είναι πολύ εύκολο ή πολύ δύσκολο γι αυτόν, συνεπώς δε μπορεί προσφέρει κάποια ένδειξη για την ικανότητά του.
Υπάρχουν δύο εναλλακτικές λύσεις για την αντιμετώπιση αυτού του προβλήματος. Η μια είναι να μη γίνεται εκτίμηση της ικανότητας, αλλά απλά να αναφέρεται ότι ο συγκεκριμένος εξεταζόμενος έχει απαντήσει σε όλες σωστά ή σε όλες λάθος. Μια άλλη λύση είναι να χρησιμοποιηθεί μια άλλη μέθοδος για την εκτίμηση. Π.χ. μπορεί να χρησιμοποιηθεί κάποια μέθοδος που βασίζεται σε μοντέλα Bayes.

Στα γνωστά προβλήματα της μεθόδου επίσης είναι ότι οι εκτιμήσεις των ικανοτήτων συχνά παίρνουν πολύ μεγάλες τιμές. Ορισμένες υλοποιήσεις δεσμεύουν τον παράγοντα διόρθωσης στην εκτέλεση της μεθόδου Newton-Raphson για να αντιμετωπίσουν αυτό το πρόβλημα. Η τεχνική αυτή όμως δεν έχει κάποιο μαθηματικό υπόβαθρο και γενικά οδηγεί σε εσφαλμένες εκτιμήσεις, για αυτό το λόγο δε χρησιμοποιείται συχνά τα τελευταία χρόνια.

Ένα άλλο γνωστό πρόβλημα της μεθόδου είναι ότι σε ορισμένες περιπτώσεις, όταν χρησιμοποιείται το μοντέλο τριών παραμέτρων, η συνάρτηση της πιθανοφάνειας παρουσιάζει περισσότερα του ενός σημεία καμπής. Το γεγονός αυτό κάνει την εκτίμηση των παραμέτρων αρκετά δύσκολη, καθώς απαιτούνται περισσότεροι και πολυπλοκότεροι υπολογισμοί. Αυτό συμβαίνει γιατί η μέθοδος Newton-Raphson θα δώσει το πλησιέστερο σημείο καμπής από το σημείο αρχικοποίησής της. Στην περίπτωση που υπάρχουν περισσότερα σημεία καμπής λοιπόν, πρέπει να εκτιμηθούν οι παράμετροι αρχικοποιώντας τη μέθοδο σε περισσότερα του ενός σημεία και να συγκριθούν τα αποτελέσματα.

Β.3.2 Ανεξαρτησία εκτίμησης ικανότητας από τα αντικείμενα

Ένα σημαντικό χαρακτηριστικό γνώρισμα της εκτίμησης της ικανότητας ενός εξεταζομένου χρησιμοποιώντας τη διαδικασία που αναλύθηκε στην προηγούμενη παράγραφο είναι ότι η εκτίμηση αυτή είναι ανεξάρτητη από τα αντικείμενα του διαγωνίσματος τα οποία χρησιμοποιούνται για να την καθορίσουν. Αυτό ισχύει με την προϋπόθεση ότι όλα τα αντικείμενα χρησιμοποιούνται για να μετρήσουν το ίδιο χαρακτηριστικό γνώρισμα (δηλ. την ίδια ικανότητα) και ότι οι τιμές των παραμέτρων των αντικειμένων μετρώνται με βάση μια κοινή κλίμακα.
Σύμφωνα με την ιδιότητα αυτή, αν ένα άτομο εξεταστεί σε δύο διαγωνίσματα η εκτιμούμενη ικανότητα του θα είναι η ίδια ανεξάρτητα με τις δυσκολίες των ερωτήσεων των διαγωνισμάτων. Αν για παράδειγμα το πρώτο διαγώνισμα έχει ερωτήσεις με μέση δυσκολία 0 και το δεύτερο με μέση δυσκολία 2 η ικανότητα του ατόμου θα εκτιμηθεί και από τα δύο διαγωνίσματα στο ίδιο επίπεδο. Αν η ικανότητα του ατόμου εκτιμηθεί στο 0 τότε πιθανότατα στο ευκολότερο διαγώνισμα να απαντήσει σε περισσότερες ερωτήσεις σωστά από ότι στο δεύτερο. Η εκτίμηση της ικανότητάς του όμως και στις δύο περιπτώσεις θα είναι η ίδια. Αυτό οφείλεται στο ότι η καμπύλη του αντικειμένου έχει πεδίο τιμών ικανότητας που διευρύνεται σε όλη τη μετρήσιμη κλίμακα. Έτσι τόσο τα αντικείμενα μεγάλης δυσκολίας όσο και αυτά μικρής δυσκολίας θα έχουν κάποιο σημείο στον άξονα των ικανοτήτων που αντιστοιχεί στην ικανότητα του συγκεκριμένου εξεταζόμενου. Αυτό που αλλάζει στις δύο περιπτώσεις είναι το σφάλμα της εκτίμησης. Στο παραπάνω παράδειγμα αν η εκτίμηση της ικανότητας του ατόμου είναι 0, το σφάλμα στο διαγώνισμα με μέση δυσκολία 0 θα είναι με μεγάλη πιθανότητα μικρότερο από ότι στο άλλο διαγώνισμα.

Η ανεξαρτησία της εκτίμησης της ικανότητας ενός εξεταζόμενου από τα αντικείμενα που χρησιμοποιούνται για την εκτίμηση αυτή είναι ένα από τα κύρια χαρακτηριστικά της IRT που την κάνουν να υπερέχει έναντι της κλασικής θεωρίας βαθμολόγησης. Χρησιμοποιώντας την κλασική θεωρία βαθμολόγησης, η εκτίμηση της ικανότητας κάποιου ατόμου εξαρτάται σε μεγάλο βαθμό από τη δυσκολία των ερωτήσεων που καλείται να απαντήσει. Αν κληθεί

να απαντήσει σε ένα σύνολο ερωτήσεων μεγάλης δυσκολίας θα απαντήσει σωστά σε λιγότερες από ότι αν κληθεί να απαντήσει σε ένα σύνολο ευκολότερων ερωτήσεων. Έτσι η εκτίμηση της ικανότητάς του σύμφωνα με την κλασική θεωρία βαθμολόγησης θα είναι διαφορετική στις δύο περιπτώσεις και προφανώς εξαρτάται από τη δυσκολία των ερωτήσεων.

Β.4 Εκτίμηση των παραμέτρων των αντικειμένων(Item parameter Estimation)

Πριν τη διεξαγωγή κάποιας εξέτασης με τη μέθοδο CAT είναι απαραίτητο οι ειδικοί να διεξάγουν κάποιες πειραματικές εξετάσεις, ή να αντλήσουν τα απαραίτητα στοιχεία από προηγούμενες εξετάσεις όταν αυτό είναι δυνατό. Οι εξετάσεις αυτές έχουν ερωτήσεις πολλαπλών εναλλακτικών απαντήσεων. Η ανάλυση των αποτελεσμάτων αυτών των εξετάσεων δίνει τις εκτιμήσεις για τις παραμέτρους των αντικειμένων που είναι υποψήφια προς χρήση στην τελική προσαρμοστική εξέταση, καθώς και άλλες τεχνικές πληροφορίες και στατιστικά στοιχεία σχετικά με τα αντικείμενα που βοηθούν στη σωστότερη χρήση τους σε μια εξέταση.

Για την εκτίμηση των παραμέτρων των αντικειμένων, διεξάγονται λοιπόν κάποιες πειραματικές εξετάσεις που περιλαμβάνουν τις ερωτήσεις των οποίων θέλουμε να εκτιμήσουμε τις παραμέτρους. Μεγάλος αριθμός τυχαία επιλεγμένων ατόμων καλείται να απαντήσει στις ερωτήσεις αυτές. Στην περίπτωση αυτή δεν είναι γνωστές ούτε οι παράμετροι των ερωτήσεων, ούτε και οι ικανότητα του κάθε εξεταζόμενου. Ο μεγάλος αριθμός όμως των εξεταζόμενων μπορεί να μας εξασφαλίσει μια κατανομή Gauss των ικανοτήτων των εξεταζόμενων.

Για την εκτίμηση των παραμέτρων των αντικειμένων (a,b,c) μπορούν να χρησιμοποιηθούν διάφορες μέθοδοι. Αυτές που χρησιμοποιούνται συχνότερα είναι οι joint maximum likelihood estimation, conditional maximum likelihood estimation, marginal maximum likelihood estimation, expectation-maximization estimation και μέθοδοι που χρησιμοποιούν μοντέλα Bayes θεωρώντας αυθαίρετα κάποιες αρχικές κατανομές για τις παραμέτρους.

Για την υλοποίηση της εργασίας επιλέχθηκε η Joint maximum likelihood estimation (JMLE ή από κοινού εκτίμηση μέγιστης πιθανοφάνειας) η οποία αναλύεται στην επόμενη παράγραφο.

Β.4.1 Joint Maximum Likelihood Estimation(JMLE)

Η JMLE βασίζεται στην εκτίμηση μέγιστης πιθανοφάνειας (MLE) που χρησιμοποιήθηκε για την εκτίμηση της ικανότητας ενός εξεταζόμενου. Στηρίζεται σε ένα παράδειγμα εκτίμησης που έδωσε το 1968 ο Birnbaum. Πολλά προγράμματα που χρησιμοποιούν την IRT βασίζονται σε αυτό το παράδειγμα ή κάποια παραλλαγή του.

Ο Fisher(1981) εξέτασε τις συνθήκες σύγκλισης αυτής της μεθόδου για το μοντέλο μιας παραμέτρου και βρήκε πως όταν ο αριθμός των εξεταζόμενων και των ερωτήσεων είναι

μεγάλος τότε η μέθοδος συνήθως συγκλίνει. Ο Samejima (1973) έδειξε πως διαγωνίσματα με λίγες ερωτήσεις(2-3) είναι πολύ πιθανό, όταν χρησιμοποιείται το μοντέλο τριών παραμέτρων, η πιθανοφάνεια να μην έχει μοναδικό μέγιστο, αλλά να υπάρχουν περισσότερα του ενός τοπικά μέγιστα. Ο Lord (1980) όμως έδειξε ότι όταν υπάρχουν αρκετά αντικείμενα (περισσότερα από 20) τότε η ύπαρξη περισσότερων του ενός μεγίστων δεν δημιουργεί πρόβλημα.

Η JMLE είναι μια διαδικασία δύο βημάτων η οποία εκτελείται επαναληπτικά. Στο πρώτο βήμα θεωρούνται γνωστές οι παράμετροι των αντικειμένων και εκτιμώνται οι ικανότητες όλων των εξεταζόμενων χρησιμοποιώντας την MLE που έχει αναλυθεί παραπάνω. Στο δεύτερο βήμα χρησιμοποιώντας τις ικανότητες που έχουν εκτιμηθεί στο πρώτο βήμα και θεωρώντας τες γνωστές, εκτιμώνται οι παράμετροι όλων των αντικειμένων. Στην επόμενη επανάληψη για το στάδιο της εκτίμησης των ικανοτήτων των εξεταζόμενων χρησιμοποιούνται οι εκτιμήσεις των παραμέτρων των αντικειμένων από την προηγούμενη επανάληψη οι οποίες θεωρούνται γνωστές. Η διαδικασία επαναλαμβάνεται μέχρι οι αλλαγές σε όλες τις παραμέτρους, ή σε μια από αυτές να είναι πολύ μικρές(κάτω από ένα ορισμένο όριο κατωφλίου) ή κάποιο από τα ενδεικτικά μεγέθη σφάλματος να είναι κάτω από ένα ορισμένο όριο κατωφλίου.

Αρχικά μπορεί να οριστεί κάποια αρχική τιμή για την ικανότητα του κάθε εξεταζόμενου, χρησιμοποιώντας π.χ. το raw score του στο διαγώνισμα, δηλαδή τον αριθμό των σωστών απαντήσεων που έχει δώσει αυτός. Μια άλλη εκδοχή είναι να οριστούν για όλα τα αντικείμενα οι ίδιες αρχικές τιμές για καθεμιά από τις παραμέτρους των αντικειμένων. Π.χ. να οριστεί α=1 και b=0 για όλα τα αντικείμενα.

Στο πρώτο βήμα των επαναλήψεων η εκτίμηση της ικανότητας του κάθε εξεταζόμενου γίνεται με τη μέθοδο μέγιστης πιθανοφάνειας (MLE) , όπως έχει αναλυθεί στην παράγραφο Β.3.1.

Στο δεύτερο βήμα των επαναλήψεων όπου προσδιορίζονται οι παράμετροι των αντικειμένων ο προσδιορισμός γίνεται χρησιμοποιώντας τη μέθοδο μέγιστης πιθανοφάνειας για την κάθε παράμετρο. Στην περίπτωση αυτή όμως χρειαζόμαστε τις τιμές των α και b που ταυτόχρονα κάνουν μέγιστη την πιθανοφάνεια.

Σε πολλές υλοποιήσεις τις μεθόδου, επειδή ο αριθμός των εξεταζόμενων είναι αρκετά μεγάλος (άνω των 500) αυτοί χωρίζονται σε κάθε επανάληψη σε ομάδες με βάση την εκτιμούμενη ικανότητά τους. Σε κάθε ομάδα αντιστοιχίζεται μια αντιπροσωπευτική τιμή ικανότητας. Π.Χ. οι εξεταζόμενοι με εκτιμούμενη ικανότητα -3 ως -2 θεωρούνται μια ομάδα με γενική ικανότητα -2,5 , αυτοί με ικανότητες -2 ως -1 μια άλλη με ικανότητα -1,5 κτλ. Αυτό γίνεται γιατί έχοντας πολλούς εξεταζόμενους και πολλές ερωτήσεις ο αριθμός των πράξεων που πρέπει να γίνουν είναι πάρα πολλές. Χρησιμοποιώντας ομάδες εξεταζόμενων όμως οι εκτιμήσεις γίνονται πολύ πιο γρήγορα, εισάγεται όμως και ένα σφάλμα. Γι αυτό το λόγο, και επειδή ο αριθμός των εξεταζόμενων θεώρησα ότι δε θα είναι τόσο μεγάλος δεν χρησιμοποιήθηκε ομαδοποίηση εξεταζόμενων στην εκτίμηση των παραμέτρων των αντικειμένων στην υλοποίηση του περιβάλλοντος αξιολόγησης.

Με μερική διαφόριση της εξίσωσης πιθανοφάνειας (Σ.10) ως προς α και b προκύπτουν οι αντίστοιχες μερικές παράγωγοι:

$$\frac{\partial \ln\left(L(\Theta, a, b|X)\right)}{\partial a} = \sum_{i=1}^{N}\sum_{j=1}^{M}(\theta_j - b_i)(x_{ij} - P_{ij}) \quad (\Sigma.15)$$

$$\frac{\partial \ln\left(L(\Theta, a, b|X)\right)}{\partial b} = \sum_{i=1}^{N}\sum_{j=1}^{M} a_i(P_{ij} - x_{ij}) \quad (\Sigma.16)$$

Για τη μεγιστοποίηση της πιθανοφάνειας πρέπει να μηδενίζονται ταυτόχρονα και οι δύο μερικές παράγωγοι (Σ.15 και Σ.16). Αρκεί δηλαδή να υπολογιστούν οι τιμές των a,b που ταυτόχρονα μηδενίζουν τις δύο εξισώσεις. Για την εύρεση αυτών των ριζών μπορεί να χρησιμοποιηθεί η μέθοδος Newton-Raphson για εξισώσεις πολλών μεταβλητών. Η μέθοδος αυτή για την εξίσωση δύο μεταβλητών γράφεται:

$$\begin{bmatrix} b_i \\ a_i \end{bmatrix}_{t+1} = \begin{bmatrix} b_i \\ a_i \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 lnL}{\partial {b_i}^2} & \frac{\partial^2 lnL}{\partial b_i \partial a_i} \\ \frac{\partial^2 \ln L}{\partial a_i \partial b_i} & \frac{\partial^2 lnL}{\partial {a_i}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial lnL}{\partial b_i} \\ \frac{\partial lnL}{\partial b_i} \end{bmatrix} \quad (\Sigma.17)$$

Όπου

$$A = \frac{\partial^2 lnL}{\partial {b_i}^2} = \sum_{i=1}^{N}\sum_{j=1}^{M} -{a_i}^2 P_{ij} Q_{ij}$$

$$\Delta = \frac{\partial^2 lnL}{\partial {a_i}^2} = \sum_{i=1}^{N}\sum_{j=1}^{M} -(b_i - \theta_j)^2 P_{ij} Q_{ij}$$

$$\frac{\partial^2 \ln L}{\partial a_i \partial b_i} = B = \frac{\partial^2 lnL}{\partial b_i \partial a_i} = \sum_{i=1}^{N}\sum_{j=1}^{M} a_i(\theta_j - b_i) P_{ij} Q_{ij} + P_{ij} - x_{ij}$$

Επίσης για τον 2Χ2 πίνακα ισχύει

$$\Pi^{-1} = \begin{bmatrix} A & B \\ \Gamma & \Delta \end{bmatrix}^{-1} = \frac{1}{A\Delta - \Gamma B}\begin{bmatrix} \Delta & -B \\ -\Gamma & A \end{bmatrix}$$

Αντικαθιστώντας στην Σ.17 παίρνω τις εξισώσεις για την επαναληπτική μέθοδο που θα δώσει τις παραμέτρους α, b:

$$(b_i)_{t+1} = (b_i)_t - \frac{1}{A\Delta - B^2}(\Delta E - BZ) \quad (\Sigma.18)$$

$$(\alpha_i)_{t+1} = (\alpha_i)_t - \frac{1}{A\Delta - B^2}(AZ - BE)$$

Όπου

$$E = \frac{\partial lnL}{\partial b_i} \; και \; Z = \frac{\partial lnL}{\partial b_i}$$

οι οποίες δίνονται από τις Σ.15 και Σ.16 αντίστοιχα

Τα μειονεκτήματα της MLE και τα γνωστά προβλήματά της ισχύουν και για την JMLE. Η ικανότητα δηλαδή ενός εξεταζόμενου που έχει απαντήσει όλες τις ερωτήσεις σωστά ή όλες λάθος θα εκτιμάται $+\infty$ ή $-\infty$ αντίστοιχα και αυτό θα επηρεάσει τα αποτελέσματα. Οι απαντήσεις των εξεταζόμενων αυτών θα πρέπει να εξαιρούνται από τη διαδικασία εκτίμησης παραμέτρων των αντικειμένων.

Αντίστοιχα αν μια ερώτηση την έχουν απαντήσει όλοι οι εξεταζόμενοι σωστά ή όλοι λάθος κάποιοι παράγοντες θα μηδενίζονται με αποτέλεσμα να έχουμε πολύ μεγάλες ή πολύ μικρές τιμές για τις παραμέτρους. Οι παράμετροι σε αυτές τις περιπτώσεις θεωρητικά είναι $+\infty$ ή $-\infty$ και ουσιαστικά δεν μπορεί να γίνει κάποια εκτίμηση για τις ερωτήσεις αυτές. Τέτοιες ερωτήσεις καλό θα ήταν να σημειώνονται και να μη συμπεριλαμβάνονται σε επόμενα διαγωνίσματα. Ένα άλλο πρόβλημα που μπορεί να προκύψει είναι να εκτιμηθεί το $\alpha<0$. Αυτό πρακτικά συμβαίνει στην περίπτωση που σε κάποια ερώτηση εξεταζόμενοι χαμηλότερης ικανότητας έχουν απαντήσει σωστά ενώ άλλοι υψηλότερης ικανότητας έχουν απαντήσει λάθος. Αν χρησιμοποιηθεί μια τέτοια ερώτηση σε ένα διαγώνισμα CAT τότε διαγωνιζόμενοι χαμηλότερης ικανότητας θα έχουν μεγαλύτερη πιθανότητα να απαντήσουν σωστά από άλλους υψηλότερης ικανότητας. Για να αποφευχθεί μια τέτοια κατάσταση και αυτές οι ερωτήσεις πρέπει να εξαιρούνται από τα διαγωνίσματα CAT. Στην περίπτωση που το α για μια ερώτηση είναι θετικό αλλά σχετικά μικρό π.χ. α=0,2 , για την ερώτηση αυτή οι όσο οι εξεταζόμενοι χαμηλής ικανότητας, τόσο και αυτοί υψηλής ικανότητας, θα έχουν περίπου την ίδια πιθανότητα να απαντήσουν σωστά. Τέτοιες ερωτήσεις δεν έχουν μεγάλη διακριτική ικανότητα και συνεπώς δεν προσφέρουν μεγάλη πληροφορία σε ένα διαγώνισμα CAT. Πρέπει λοιπόν να καθοριστεί ένα κατώφλι για το α , έτσι ώστε ερωτήσεις που έχουν α μικρότερο από αυτό το κατώφλι να εξαιρούνται από τα διαγωνίσματα.

Ένα μεγάλο πρόβλημα της μεθόδου είναι οι περιπτώσεις Heywood που αναλύονται στην επόμενη παράγραφο.

Όταν βρεθούν εκτιμήσεις με τη μέθοδο μέγιστης πιθανοφάνειας αυτές είναι συνεπής. Ο Birnbaum (1968) αναφέρει ότι δεν υπάρχουν επαρκή στατιστικά στοιχεία για την εκτίμηση ευσταθών παραμέτρων για το μοντέλο τριών παραμέτρων.

Σχετικά με τη συνέπεια των εκτιμήσεων των παραμέτρων, ο Andersen (1973) απέδειξε ότι χρησιμοποιώντας το μοντέλο μιας παραμέτρου, οι εκτιμήσεις της μεθόδου μέγιστης πιθανοφάνειας δεν είναι συνεπείς όταν το διαγώνισμα έχει σταθερό αριθμό ερωτήσεων και ο αριθμός των εξεταζόμενων πλησιάζει το άπειρο. Ο Haberman(1977) έδειξε ότι αν αυξάνεται και ο αριθμός των ερωτήσεων παράλληλα με αυτόν των εξεταζόμενων τότε οι εκτιμήσεις των παραμέτρων δυσκολίας είναι συνεπείς. Γενικά για την ευστάθεια των αποτελεσμάτων ο αριθμός των εξεταζόμενων και αυτός των ερωτήσεων πρέπει να έχουν

ένα συγκεκριμένο λόγο. Όταν αυξάνει ο αριθμός των εξεταζόμενων πρέπει ανάλογα να αυξάνει και ο αριθμός των ερωτήσεων.

Σε γενικές γραμμές η JMLE είναι ευσταθής για το μοντέλο 1PL ακόμα και αν ο αριθμός δειγμάτων που υπάρχουν είναι μικρός. Για τα μοντέλα 2PL και 3PL όμως μπορεί να μη συγκλίνει. Επίσης στα μοντέλα αυτά παρουσιάζονται τοπικά μέγιστα της συνάρτησης ln(L) όταν ο αριθμός των δειγμάτων είναι μικρός. Σε τέτοιες περιπτώσεις ακόμη και αν συγκλίνει η μέθοδος η τελική εκτίμηση εξαρτάται από το σημείο αρχικοποίησης. Για διαφορετικά δηλαδή σημεία αρχικοποίησης μπορεί να συγκλίνει σε διαφορετικές λύσεις όταν παρουσιάζονται τοπικά μέγιστα. Ένα άλλο πρόβλημα που μπορεί κανείς να συναντήσει είναι να μη συγκλίνει η μέθοδος Newton-Raphson σε κάποιο βήμα αλλά να παλινδρομεί μεταξύ δύο τιμών. Αυτό συμβαίνει κυρίως στην περίπτωση που η ln(L) δεν παρουσιάζει εμφανές μέγιστο, αλλά παίρνει τη μέγιστη τιμή της για μια περιοχή των (a,b). Λόγω αυτών των δυσκολιών πολλές φορές χρησιμοποιούνται άλλες μέθοδοι όπως η E-M. Η JMLE προσπαθεί να εκτιμήσει τις παραμέτρους με ελλειπή δεδομένα εισόδου, καθώς οι ικανότητες των εξεταζόμενων δεν είναι εκ των προτέρων γνωστές, αλλά εκτιμούνται κατά την εκτέλεση της μεθόδου. Εναλλακτικές μέθοδοι για την εκτίμηση των παραμέτρων όπως η E-M προσπαθούν να εκτιμήσουν της παραμέτρους θεωρώντας γνωστές τις ικανότητες των εξεταζόμενων. Για των προσδιορισμό των ικανοτήτων αυτών θεωρείται ότι αυτές είναι κατανεμημένες σύμφωνα με μια κατανομή που ορίζεται εκ των προτέρων.

Β.4.2 Περιπτώσεις Heywood (Heywood cases)

Ένα μεγάλο πρόβλημα της μεθόδου είναι ότι σε ορισμένα σετ αποτελεσμάτων για ένα ή περισσότερα αντικείμενα ο παράγοντας διακριτικής ικανότητας μπορεί να πάρει μεγάλες τιμές. Αυτό οδηγεί σε μεγάλες τιμές εκτιμούμενης ικανότητας για τους εξεταζόμενους που απάντησαν σωστά σε αυτά τα αντικείμενα. Έτσι, επειδή η μέθοδος είναι επαναληπτική, σταδιακά οι εκτιμήσεις του παράγοντα διακριτικής ικανότητας τείνουν στο άπειρο. Το πρόβλημα αυτό συναντάται όταν χρησιμοποιείται το μοντέλο δύο ή το μοντέλο τριών παραμέτρων, αλλά όχι όταν χρησιμοποιείται το μοντέλο μιας παραμέτρου στο οποίο η παράμετρος διακριτικής ικανότητας θεωρείται σταθερή και ίση με ένα για όλα τα αντικείμενα.

Το φαινόμενο αυτό είναι γνωστό σαν περίπτωση Heywood (Heywood case) και συναντάται στην ανάλυση παραγόντων. Μια περίπτωση Heywood προκύπτει όταν για μια ή περισσότερες εκτιμήσεις παραγόντων η τιμή της διασποράς του σφάλματος είναι μηδέν ή αρνητική (Chatfield and Collins, 1980, p.87). Στην JMLE όταν συμβαίνει αυτό πρακτικά μια ή περισσότερες παράμετροι διακριτικής ικανότητας γίνονται άπειρο. Οι περιπτώσεις Heywood είναι ένα πρόβλημα που προκύπτει συχνά όταν χρησιμοποιούνται μέθοδοι μέγιστης πιθανοφάνειας , σύμφωνα με το Bartholomew et al., 2002, (παρ. 2.5.2.1)

Σύμφωνα με τις οδηγίες χρήσης του προγράμματος ανάλυσης στατιστικών στοιχείων SAS/STAT οι περιπτώσεις Heywood πιθανώς να οφείλονται σε

- Μικρό αριθμό δειγμάτων, ο οποίος δεν αρκεί για να εξασφαλίσει ευσταθείς εκτιμήσεις

- Πολύ μικρό αριθμό κοινών παραγόντων (common factors)
- Πολύ μεγάλο αριθμό κοινών παραγόντων (common factors)
- Το μοντέλο που χρησιμοποιείται δεν είναι κατάλληλο για τα δεδομένα που αναλύονται

Το πρόβλημα αντιμετωπίζεται περιορίζοντας τη διασπορά των σφαλμάτων σε τιμές μεγαλύτερες από κάποιο θετικό αριθμό ή εφόσον χρησιμοποιείται η μέθοδος της μέγιστης πιθανοφάνειας, αυτή να αρχικοποιείται σε τιμές κατάλληλες για το εκάστοτε σύνολο δειγμάτων. Οι Tucker and MacCallum (1997, p.266-282) προτείνουν μια εναλλακτική βαθμιδωτή μέθοδο.

Το πρόβλημα αυτό μπορεί επίσης να αντιμετωπισθεί χρησιμοποιώντας μεθόδους που βασίζονται σε μοντέλα Bayes αντί για τη μέθοδο μέγιστης πιθανοφάνειας . Οι μέθοδοι αυτοί όμως απαιτούν από το χρήστη να υποθέσει εξ αρχής κάποιες κατανομές για τις παραμέτρους και αυτό δεν έχει διαπιστωθεί αν όντως οδηγεί πάντα σε σωστά αποτελέσματα.

Β.4.3 Σφάλμα εκτιμήσεων

Αφού εκτιμηθούν οι παράμετροι των αντικειμένων, για κάθε αντικείμενο το σύνολο των εκτιμούμενων παραμέτρων του ορίζουν μια καμπύλη αντικειμένου. Θα ήταν χρήσιμο να γνωρίζουμε το πόσο ταιριάζει η καμπύλη αυτή με τα πειραματικά δεδομένα, δηλ. το σύνολο των απαντήσεων των εξεταζόμενων στο συγκεκριμένο αντικείμενο. Χρησιμοποιώντας την JMLE ένα μέγεθος που δίνει μια ένδειξη για το πόσο συμφωνούν τα πειραματικά δεδομένα με τις εκτιμούμενες παραμέτρους είναι ο δείκτης σφάλματος chi-square.

Στην περίπτωση που χρησιμοποιείται ομαδοποίηση των εξεταζομένων κατά την εκτίμηση των παραμέτρων ο δείκτης αυτός δίνεται από τον τύπο:

$$x^2 = \sum_{j=1}^{J} m_j \frac{[p(\theta_j) - P(\theta_j)]^2}{P(\theta_j)Q(\theta_j)} \quad (\Sigma. 19)$$

Όπου J είναι ο αριθμός των ομάδων που έχουν χωριστεί οι εξεταζόμενοι σύμφωνα με την ικανότητά τους

θ_j είναι η ικανότητα που αντιστοιχεί στην ομάδα j

m_j είναι ο αριθμός των εξεταζόμενων που ανήκουν στην ομάδα j

$p(\theta_j)$ είναι η παρατηρούμενη πιθανότητα σωστής απάντησης για την ομάδα j. Αυτή ισούται με $p(\theta_j) = \frac{r_j}{m_j}$

r_j είναι ο αριθμός των ατόμων στην ομάδα j που έχει απαντήσει σωστά στο συγκεκριμένο αντικείμενο.

$P(\theta_j)$ είναι η πιθανότητα σωστής απάντησης για την ομάδα j που δίνεται από τη χαρακτηριστική του αντικειμένου χρησιμοποιώντας τις παραμέτρους που έχουν εκτιμηθεί.

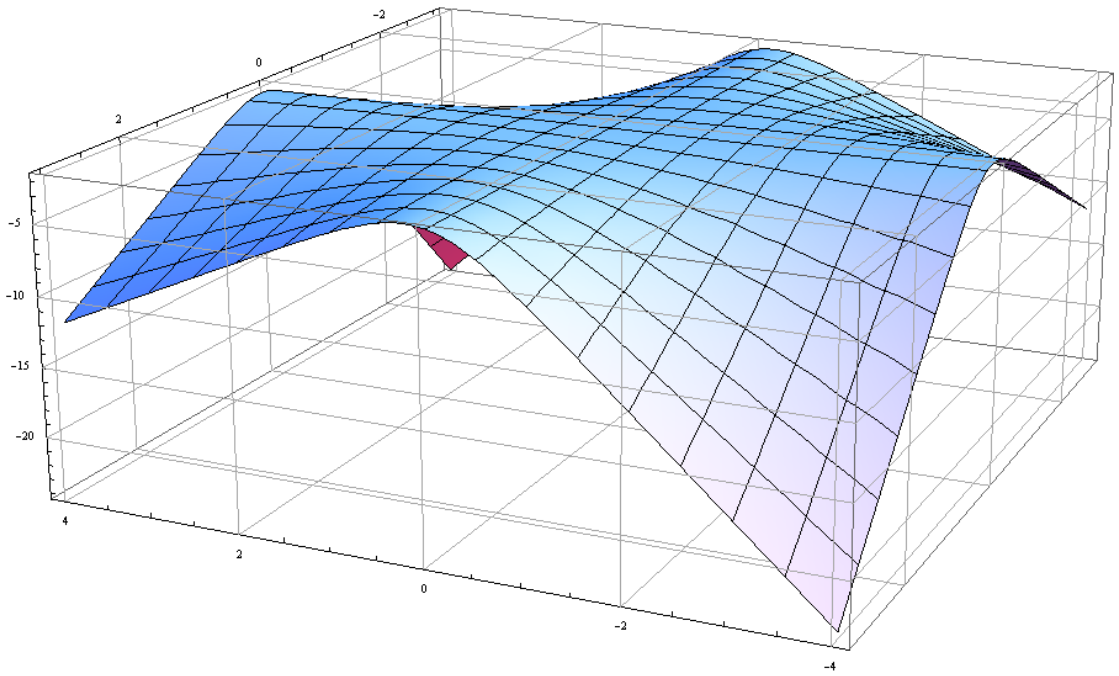
Στην περίπτωση που δεν χρησιμοποιείται κάποια ομαδοποίηση των εξεταζόμενων κατά την εκτίμηση των παραμέτρων ο παραπάνω δείκτης σφάλματος γράφεται:

$$x^2 = \sum_{j=1}^{J} \frac{[u_j - P(\theta_j)]^2}{P(\theta_j)Q(\theta_j)} \quad (\Sigma.20)$$

Όπου u_jη βαθμολογία του εξεταζόμενου j, η οποία ισούται με 0 ή 1 ανάλογα με το αν έχει απαντήσει λανθασμένα ή σωστά στο συγκεκριμένο αντικείμενο.

Ο παραπάνω δείκτης είναι δείκτης σφάλματος, δηλαδή όσο μικρότερος είναι ο δείκτης τόσο καλύτερα συμφωνούν τα πειραματικά δεδομένα από το σύνολο των απαντήσεων με τις εκτιμούμενες τιμές των παραμέτρων.

Για μεγάλο αριθμό δειγμάτων (πάνω από 1000) τα στατιστικά σφάλματα chi-square παίρνουν μεγάλες τιμές. Το γεγονός αυτό οφείλεται στον αριθμό των δειγμάτων και όχι στο πόσο καλά ταιριάζουν οι εκτιμήσεις με τα δεδομένα των δειγμάτων. Γι αυτό, όταν υπάρχει μεγάλος αριθμός δειγμάτων πρέπει αυτό να λαμβάνεται υπ' όψη και η τιμή του σφάλματος chi-square ενός αντικειμένου να εκτιμάται σε σχέση με τις αντίστοιχες τιμές των υπολοίπων αντικειμένων.
Κατά την ανάλυση των πειραματικών δεδομένων συνήθως ορίζεται από τον αναλυτή, ανάλογα με τα δεδομένα της εξέτασης, ένα όριο για την τιμή του σφάλματος chi-square. Όταν η τιμή σφάλματος ενός αντικειμένου ξεπεράσει το όριο αυτό θεωρείται ότι για το αντικείμενο αυτό οι εκτιμούμενες παράμετροι δίνουν μια καμπύλη που δεν συμφωνεί με τα πειραματικά δεδομένα και το αντικείμενο απορρίπτεται. Αυτό μπορεί να συμβεί εάν οι σωστές απαντήσεις για το αντικείμενο αυτό είναι διεσπαρμένες τόσο πολύ σε διάφορα επίπεδα ικανότητας των εξεταζόμενων που δεν είναι δυνατό να βρεθεί κάποιο σύνολο παραμέτρων που να συμφωνεί με τα δεδομένα αυτά, ανεξάρτητα με το μοντέλο που χρησιμοποιείται. Γι αυτό το λόγο και όταν υπάρχουν λίγα αντικείμενα τα σφάλματα είναι πολύ μεγάλα. Μια άλλη περίπτωση είναι να χρησιμοποιείται λάθος μοντέλο για τα συγκεκριμένα δεδομένα. Προφανώς κάποιο από τα μοντέλα θα συμφωνεί περισσότερο με τα δεδομένα και συνεπώς το σφάλμα χρησιμοποιώντας αυτό το μοντέλο θα είναι μικρότερο από ότι στα άλλα για τα ίδια δεδομένα.



Σχήμα 9 : *Η συνάρτηση* $\sum_{j=1}^{3} lnL$ *για τρία αντικείμενα.*

Στο σχήμα 9 φαίνεται η γραφική παράσταση της συνάρτησης $\sum_{j=1}^{3} lnL$ ως προς b[-4,4] και α[-3,3] για τρία αντικείμενα του μοντέλου 2PL που αναφέρονται σε μια ερώτηση. Όπως φαίνεται και από το σχήμα η συνάρτηση δεν έχει μοναδικό σημείο (a,b) στο οποίο μεγιστοποιείται. Οποιαδήποτε από τα σημεία γύρω από το (0,0) είναι αποδεκτές λύσεις καθώς μεγιστοποιούν τη συνάρτηση. Επειδή η JMLE είναι επαναληπτική μέθοδος τέτοιες περιπτώσεις μπορεί να οδηγήσουν σε μεγάλα σφάλματα εκτιμήσεων. Επίσης τα σφάλματα που υπολογίζονται για τις εκτιμήσεις της JMLE είναι σχετικά, καθώς δεν λαμβάνουν υπ όψη τα σφάλματα εκτίμησης των θ που εισέρχονται στο ένα από τα δύο βήματα της μεθόδου.

Β.5 Συναρτήσεις πληροφορίας

Το μέγεθος που χρησιμοποιείται από την IRT για τη μέτρηση της πληροφορίας που περιέχει ένα αντικείμενο σχετικά με την άγνωστη παράμετρο της ικανότητας ενός εξεταζομένου είναι η **πληροφορία Fisher**. Το μέγεθος αυτό χρησιμοποιείται συχνά στη στατιστική και τη θεωρία της πληροφορίας, συμβολίζεται με I(θ) και πήρε το όνομα του Άγγλου μαθηματικού και στατιστικού αναλυτή sir Roland Aylmer Fisher, ο οποίος το εισήγαγε. Σύμφωνα με τον ορισμό του Fisher η πληροφορία είναι μέγεθος ανάλογο της ακρίβειας με την οποία εκτιμάται μια παράμετρος. Πρακτικά αυτό σημαίνει ότι όταν μια παράμετρος εκτιμηθεί με μεγάλη ακρίβεια υπάρχει και μεγάλη πληροφορία σχετικά με αυτή την παράμετρο. Το μέγεθος που μετρά την ακρίβεια με την οποία έχει εκτιμηθεί μια παράμετρος είναι η

διασπορά σ^2 των εκτιμούμενων τιμών. Έτσι το μέγεθος της πληροφορίας Fisher δίνεται από τη σχέση:

$$I = \frac{1}{\sigma^2} \quad (\Sigma. 21)$$

Στην παράγραφο Β.3.1 έχει υπολογιστεί το τυπικό σφάλμα της εκτίμησης της παραμέτρου θ για τα μοντέλα μιας και δύο παραμέτρων, η οποία δίνεται από την εξίσωση :

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^{N} a^2{}_i P(\hat{\theta}) Q(\hat{\theta})}} \qquad (\Sigma. 22)$$

Αυτό είναι ένα μέτρο της διασποράς των εκτιμούμενων τιμών της θ. Από τους δύο παραπάνω τύπους προκύπτει ότι:

$$I(\theta) = \sum_{i=1}^{N} a^2{}_i P(\hat{\theta}) Q(\hat{\theta}) \qquad (\Sigma. 23)$$

Αυτή είναι η πληροφορία που έχουμε σχετικά με την παράμετρο θ η οποία έχει εκτιμηθεί από τα N αντικείμενα i=1….N. Η συνάρτηση πληροφορίας Σ.23 αναφέρεται συνήθως σαν **συνάρτηση πληροφορίας διαγωνίσματος (test information function)**

Μια πολύ χρήσιμη ιδιότητα της πληροφορίας Fisher είναι ότι είναι αθροιστική. Δηλαδή αν υποθέσουμε ότι έχουμε δύο ανεξάρτητα τυχαία πειράματα, η πληροφορία που αντλείται από τα δύο αυτά πειράματα είναι το άθροισμα της πληροφορίας που θα αντλούνταν από το κάθε ένα πείραμα αν αυτό γινόταν ξεχωριστά και ανεξάρτητα από το άλλο. Όμοια αν έχουμε N ανεξάρτητα πειράματα η συνολική πληροφορία από αυτά είναι το άθροισμα των επιμέρους πληροφοριών. Για N λοιπόν αντικείμενα IRT η συνολική πληροφορία είναι το άθροισμα των επιμέρους αντικειμένων δηλ.

$$I(\theta) = I_1 + I_2 + \cdots + I_N = \sum_{i=1}^{N} I_i \qquad (\Sigma. 24)$$

Όπου I_i η πληροφορία που περιέχει το αντικείμενο i.
Από τις σχέσεις Σ.23 και Σ.24 προκύπτει η σχέση που δίνει την πληροφορία που περιέχει για την εκτίμηση της παραμέτρου ικανότητας θ το αντικείμενο i.

$$I_i(\theta) = a^2{}_i P(\hat{\theta}) Q(\hat{\theta}) \qquad (\Sigma. 25)$$

Το μέγεθος αυτό συνήθως αναφέρεται ως **πληροφορία αντικειμένου (item Information)**

Επαναλαμβάνοντας την παραπάνω διαδικασία για το μοντέλο τριών παραμέτρων, για το οποίο είχε υπολογιστεί

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^{N} a^2{}_i \frac{Q(\hat{\theta})}{P(\hat{\theta})} \left[\frac{P(\hat{\theta}) - c}{1 - c}\right]^2}}$$

Υπολογίζονται οι εξισώσεις που δίνουν την πληροφορία διαγωνίσματος και την πληροφορία αντικειμένου για το μοντέλο αυτό.

$$I_{3PL}(\theta) = \sum_{i=1}^{N} a^2{}_i \frac{Q(\hat{\theta})}{P(\hat{\theta})}\left[\frac{P(\hat{\theta}) - c}{1 - c}\right]^2$$

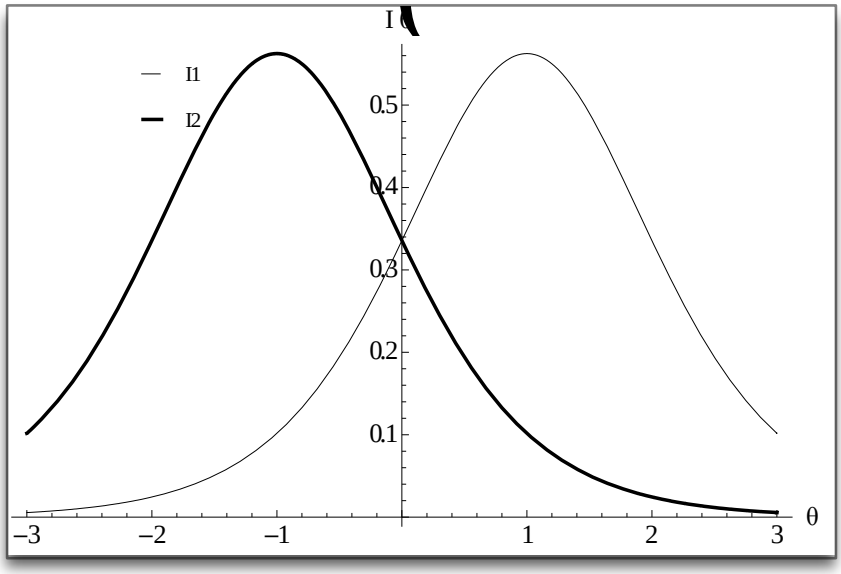
$$I_{i_{3PL}}(\theta) = a^2{}_i \frac{Q(\hat{\theta})}{P(\hat{\theta})}\left[\frac{P(\hat{\theta}) - c}{1 - c}\right]^2$$

Β.5.1 Ιδιότητες συναρτήσεων πληροφορίας

Σχεδιάζοντας τη γραφική παράσταση της πληροφορίας Fisher σαν συνάρτηση της ικανότητας μπορεί κάποιος να δει ορισμένες χρήσιμες ιδιότητες της συνάρτησης αυτής.
Στο Σχήμα 10 φαίνεται η πληροφορία Fisher δύο αντικειμένων, I1 και I2, του μοντέλου δύο παραμέτρων, με ίδιες παραμέτρους διακριτικής ικανότητας (a=1,5) και διαφορετικές παραμέτρους δυσκολίας. Συγκεκριμένα το πρώτο έχει b=+1 και το δεύτερο b=-1. Μπορεί κανείς να δει ότι η πληροφορία Fisher είναι πάντοτε μεγαλύτερη του μηδενός, όπως θα περίμενε και από τους τύπους Σ.24 και Σ.25 που δίνουν την πληροφορία.

Η πληροφορία παρουσιάζει μέγιστο για ικανότητα ίση με την τιμή της παραμέτρου δυσκολίας b του αντικειμένου. Χρησιμοποιώντας δηλαδή σε ένα διαγώνισμα ένα αντικείμενο δυσκολίας κοντά στην πραγματική ικανότητα του εξεταζομένου, το αντικείμενο αυτό περιέχει περισσότερη πληροφορία για την ικανότητα αυτή από την πληροφορία που θα περιείχε ένα άλλο με δυσκολία πολύ μεγαλύτερη ή πολύ μικρότερη από την ικανότητα του εξεταζομένου.

Για μεγάλες τιμές απόλυτης τιμής του θ η συνάρτηση πλησιάζει ασυμπτωτικά το μηδέν και πρακτικά μηδενίζεται. Αν π.χ. χρησιμοποιούνταν το αντικείμενο I1 του Σχήματος 10 για την εκτίμηση της ικανότητας ενός εξεταζόμενου ικανότητας θ= - 3 το αντικείμενο αυτό δεν μπορεί να δώσει μεγάλη πληροφορία για την ικανότητα του. Η πιθανότητα να απαντήσει σωστά σε μια ερώτηση τέτοιας δυσκολίας είναι πολύ μικρή. Αντίθετα για έναν εξεταζόμενο Ικανότητας θ=1 το αντικείμενο περιέχει μεγάλη πληροφορία.



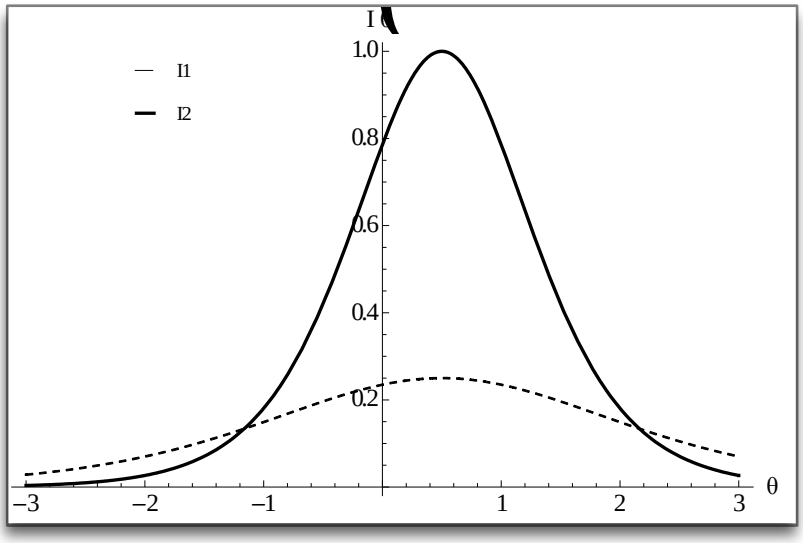
Σχήμα 10 : *Η πληροφορία Fisher για δύο αντικείμενα (I1,I2).*

Παράμετροι αντικειμένων :

Αντικείμενο	*b*	*a*
I1	*+1*	*+1.5*
I2	*-1*	*+1.5*

Αντικείμενα με δυσκολία κοντά στην πραγματική ικανότητα του εξεταζόμενου λοιπόν μπορούν να προσφέρουν μεγαλύτερη πληροφορία για την εκτίμηση της ικανότητας αυτής.

Στο Σχήμα 11 φαίνεται η πληροφορία Fisher δύο αντικειμένων, I1 και I2, τα οποία έχουν την ίδια παράμετρο δυσκολίας. Η παράμετρος διακριτικής ικανότητας του I1 όμως είναι μικρότερη από αυτή του I2. Μπορεί κανείς να δει πως η πληροφορία που περιέχει το αντικείμενο με τη μεγαλύτερη διακριτική ικανότητα είναι μεγαλύτερη στην περιοχή ικανοτήτων που πρακτικά μας ενδιαφέρει, δηλ. για θ κοντά στη δυσκολία των αντικειμένων.



Σχήμα 11 : *Η πληροφορία Fisher για δύο αντικείμενα (I1,I2).*

Παράμετροι αντικειμένων :

Αντικείμενο	*b*	*a*
I1	*+0.5*	*+1*
I2	*+0.5*	*+2*

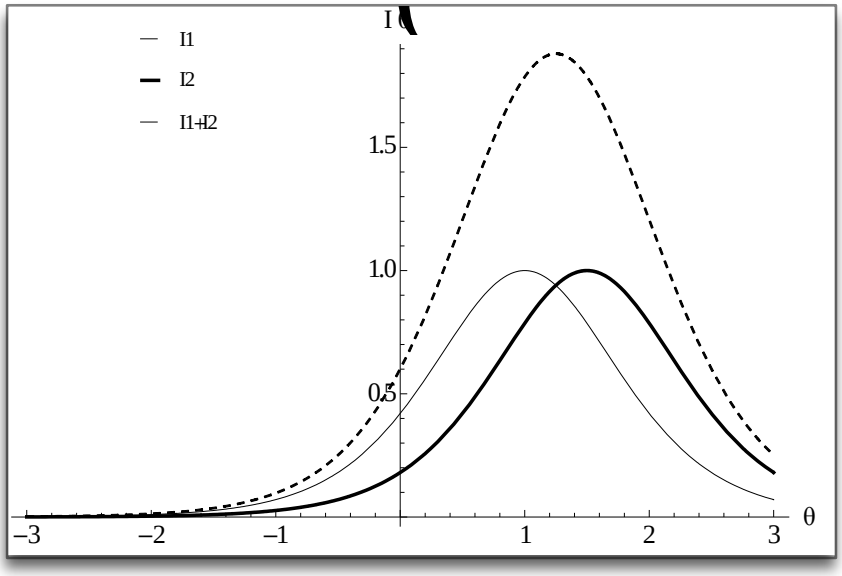
Με την προϋπόθεση ότι δύο ή περισσότερα αντικείμενα είναι ανεξάρτητα μεταξύ τους, μπορεί να χρησιμοποιηθεί η αθροιστική ιδιότητα της πληροφορίας Fisher για να βρεθεί η πληροφορία που περιέχει ένα σύνολο αντικειμένων. Στο Σχήμα 12 φαίνεται η πληροφορία που περιέχουν δύο αντικείμενα I1, I2 με παραμέτρους δυσκολίας 1 και 1,5 αντίστοιχα. Στο Σχήμα 13 φαίνεται η πληροφορία που περιέχουν δύο άλλα με παραμέτρους δυσκολίας 0,5 και 1,5 αντίστοιχα. Αν υποτεθεί ότι τα δύο σύνολα αντικειμένων αποτελούν δύο διαγωνίσματα τα οποία έχουν δύο αντικείμενα το καθένα, η μέση δυσκολία του πρώτου διαγωνίσματος είναι 1,25 ενώ του δευτέρου 1. Αυτές είναι οι τιμές δυσκολίας για τις οποίες γίνεται μέγιστη και η πληροφορία στο κάθε διαγώνισμα για τα συγκεκριμένα διαγωνίσματα. Η μέγιστη πληροφορία που περιέχει το πρώτο διαγώνισμα όπως φαίνεται και από τα σχήματα είναι μεγαλύτερη.

Χρησιμοποιώντας λοιπόν το πρώτο διαγώνισμα μπορεί να εκτιμηθεί η ικανότητα κάποιων εξεταζομένων με μεγαλύτερη ακρίβεια από ότι αν χρησιμοποιηθεί το δεύτερο. Για να ισχύει αυτό , σύμφωνα και με τα προηγούμενα συμπεράσματα πρέπει η πραγματική ικανότητα των εξεταζομένων που καλούνται να απαντήσουν στα διαγωνίσματα να βρίσκεται στο επίπεδο δυσκολίας για το οποίο γίνεται μέγιστη η πληροφορία του εκάστοτε διαγωνίσματος.

Το συμπέρασμα είναι ότι όσο μεγαλύτερη είναι η πληροφορία που έχει ένα σύνολο αντικειμένων που έχει χρησιμοποιηθεί για την εκτίμηση της ικανότητας ενός εξεταζόμενου , τόσο ακριβέστερη είναι η εκτίμηση αυτή , αφού το σφάλμα εκτίμησης ελαχιστοποιείται.

Λόγω της αθροιστικής ιδιότητας της πληροφορίας επίσης , και επειδή αυτή είναι πάντα θετική μπορεί εύκολα κάποιος να συμπεράνει ότι όσο περισσότερα αντικείμενα χρησιμοποιούνται τόσο μεγαλύτερη θα είναι η πληροφορία του συνόλου των αντικειμένων , και κατά συνέπεια τόσο μικρότερο θα είναι το σφάλμα εκτίμησης.

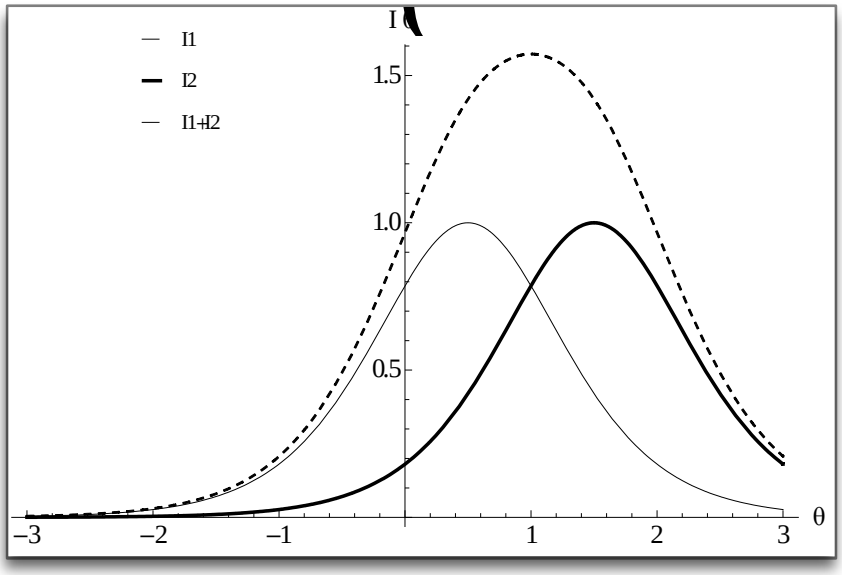
Συνοψίζοντας μπορεί να επιτευχθεί ακριβέστερη εκτίμηση της ικανότητας ενός εξεταζόμενου χρησιμοποιώντας όσο το δυνατό περισσότερα αντικείμενα. Επιπλέον η εκτίμηση είναι πιο ακριβής χρησιμοποιώντας αντικείμενα επιπέδου δυσκολίας κοντά στην τελική εκτίμηση της ικανότητας του εξεταζόμενου και όσο το δυνατό μεγαλύτερης διακριτικής ικανότητας.



Σχήμα 12 : *Η πληροφορία Fisher για δύο αντικείμενα (I1,I2). Με διακεκομμένη γραμμή φαίνεται η συνολική πληροφορία που περιέχουν τα δύο αντικείμενα.*

Παράμετροι αντικειμένων :

Αντικείμενο	*b*	*a*
I1	*+1*	*+2*
I2	*+1,5*	*+2*



Σχήμα 13 : *Η πληροφορία Fisher για δύο αντικείμενα (I1,I2).. Με διακεκομμένη γραμμή φαίνεται η συνολική πληροφορία που περιέχουν τα δύο αντικείμενα.*

Παράμετροι αντικειμένων :

Αντικείμενο	*b*	*a*
I1	*+0.5*	*+2*
I2	*+1,5*	*+2*

Γ. Προσαρμοστικές μέθοδοι αξιολόγησης μέσω Ηλεκτρονικών Υπολογιστών (Computer Adaptive Testing C.A.T)

Γ.1 Κλασική θεωρία αξιολόγησης

Σύμφωνα τον κλασικό τρόπο αξιολόγησης μέσω γραπτών εξετάσεων το διαγώνισμα που καλούνται να απαντήσουν οι εξεταζόμενοι είναι κοινό για όλους. Όπως μπορεί εύκολα να συμπεράνει κάποιος από την παράγραφο Β.5.1 , για να μπορέσει να εκτιμηθεί η ικανότητα ενός εξεταζόμενου με κάποια αποδεκτή ακρίβεια χρησιμοποιώντας αυτή τη μέθοδο , θα έπρεπε το διαγώνισμα να αποτελείται από ερωτήσεις σε όλα τα επίπεδα δυσκολίας. Διαφορετικά , αν για παράδειγμα το διαγώνισμα αποτελείται από πολλές εύκολες ερωτήσεις , η εκτίμηση θα έχει πολύ μεγάλο σφάλμα για όλους τους εξεταζόμενους που η ικανότητάς τους είναι μεγαλύτερη από αυτό το επίπεδο δυσκολίας. επιπλέον όλοι οι εξεταζόμενοι καλούνται να απαντήσουν σε όλες τις ερωτήσεις , με αποτέλεσμα κάποιοι εξεταζόμενοι υψηλής ικανότητας να καλούνται να απαντήσουν σε κάποιες πολύ εύκολες για αυτόν ερωτήσεις και αντίστοιχα κάποιοι εξεταζόμενοι χαμηλής ικανότητας καλούνται να απαντήσουν σε κάποιες ερωτήσεις που είναι πολύ δύσκολές γι αυτούς. Σύμφωνα με τη θεωρία της πληροφορίας Fisher , αυτές οι ερωτήσεις προσφέρουν ελάχιστη πληροφορία για την εκτίμηση της ικανότητας του εξεταζόμενου.

Αν για παράδειγμα κληθούν κάποιοι εξεταζόμενοι να απαντήσουν σε ένα διαγώνισμα εννέα ερωτήσεων , στο οποίο οι δυσκολίες των τριών ερωτήσεων κυμαίνονται στο επίπεδο -3 ως -1 , οι επόμενες τρεις στο επίπεδο -1 ως +1 και οι τελευταίες τρεις στο επίπεδο δυσκολίας +1 ως +3, το σφάλμα εκτίμησης για όλους τους εξεταζόμενους θα ήταν σχετικά υψηλό αφού οι περισσότερες ερωτήσεις περιέχουν ελάχιστη πληροφορία σχετικά με το επίπεδο ικανότητας. Επιπλέον το σφάλμα θα ήταν στα ίδια επίπεδα για όλους τους εξεταζόμενους , λόγω της ομοιόμορφης κατανομής των ερωτήσεων με βάση τη δυσκολία τους.

Αν στο προηγούμενο παράδειγμα υποτεθεί ότι το δείγμα των εξεταζόμενων είναι τυχαίο και οι ικανότητες τους ακολουθούν μια κατανομή Gauss , θα μπορούσε κανείς να κατανέμει με παρόμοιο τρόπο και τις ερωτήσεις στο διαγώνισμα με βάση τη δυσκολία τους. Έχοντας δηλαδή περισσότερες ερωτήσεις μέσου επιπέδου δυσκολίας θα έπαιρνε εκτιμήσεις με μικρότερο σφάλμα για τους εξεταζόμενους μέσου επιπέδου ικανότητας. Το σφάλμα όμως για αυτούς με πολύ υψηλό ή πολύ χαμηλό επίπεδο ικανότητας θα ήταν μεγαλύτερο από ότι με την ομοιόμορφη κατανομή.

Η ακρίβεια με την οποία μπορεί με τους παραπάνω τρόπους να εκτιμηθεί η ικανότητα ενός εξεταζόμενου περιορίζεται και από τις πραγματικές συνθήκες και τα όρια του ανθρώπινου οργανισμού. Για να έχει δηλαδή μια ικανοποιητική ακρίβεια (με σφάλμα εκτίμησης κάτω του 1/10) πρέπει να έχει απαντήσει σε ένα ικανοποιητικό σύνολο ερωτήσεων με δυσκολία κοντά στο επίπεδο ικανότητάς του. Δεν είναι φυσικά δυνατό όμως να διοργανώνονται εξετάσεις με διαγωνίσματα των 50 η παραπάνω ερωτήσεων. Ένα τέτοιο διαγώνισμα θα έπρεπε να διαρκεί τουλάχιστο 100 λεπτά. Η κόπωση των εξεταζόμενων μπορεί να επηρεάσει την επίδοσή τους. Επιπλέον το μεγαλύτερο μέρος των δεδομένων από ένα

τέτοιο διαγώνισμα προσφέρει από ελάχιστη ως καθόλου πληροφορία για την ικανότητα του κάθε εξεταζόμενου. Μια λύση για μην είναι τόσο κουραστική η εξέταση θα ήταν να γίνουν περισσότερες εξετάσεις με μικρότερα διαγωνίσματα. Αυτή η λύση όμως δεν λύνει το πρόβλημα ότι το μεγαλύτερο μέρος των δεδομένων δεν προσφέρουν καμία πληροφορία. Επιπλέον η διόρθωση ενός τέτοιου διαγωνίσματος απαιτεί σημαντικό χρόνο και κατάλληλα εκπαιδευμένα άτομα.

Γ.2 Προσαρμοστικές μέθοδοι αξιολόγησης μέσω ηλεκτρονικών υπολογιστών

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλά διαγωνίσματα μέσω ηλεκτρονικού υπολογιστή τα οποία βασίζονται στην IRT. Σκοπός αυτών των διαγωνισμάτων είναι η ακριβής εκτίμηση της ικανότητας των εξεταζόμενων , χρησιμοποιώντας όσο το δυνατό λιγότερα αντικείμενα. Αυτό το επιτυγχάνουν ζητώντας από τον κάθε εξεταζόμενο να απαντήσει μόνο στα απαραίτητα αντικείμενα για την ακριβή εκτίμηση της ικανότητάς του. Χρησιμοποιούνται δηλαδή για τον καθένα μόνο τα αντικείμενα που προσφέρουν κάποια πληροφορία για την ικανότητά του, τα οποία είναι διαφορετικά για τον κάθε εξεταζόμενο και εξαρτώνται από το επίπεδο ικανότητάς του. Τα διαγωνίσματα αυτά έχουν τη δυνατότητα να προσαρμόζονται στο επίπεδο του εξεταζόμενου , και γι αυτό είναι γνωστά σαν προσαρμοστικά διαγωνίσματα ή προσαρμοστικές μέθοδοι αξιολόγησης μέσω ηλεκτρονικού υπολογιστή (Computer Adaptive Tests ή C.A.T).

Επειδή τα CAT βασίζονται στην IRT μπορούν να εκτιμούν με μεγάλη ακρίβεια την ικανότητα των εξεταζόμενων χρησιμοποιώντας λίγα αντικείμενα. Η ακρίβεια όμως και η συντομότερη εξέταση δεν είναι τα μόνα σημεία στα οποία υπερέχουν τα CAT έναντι των κλασικών εξετάσεων. Μια σύγκριση των δύο μεθόδων εξέτασης γίνεται στην παράγραφο Γ.3.

Τα διαγωνίσματα CAT βασίζονται σε τράπεζες αντικειμένων τα οποία έχουν αναλυθεί πριν με βάση την IRT. Για το λόγο αυτό πριν τη διεξαγωγή ενός CAT πρέπει να προηγηθεί η κατάλληλη ανάλυση του διαγωνίσματος. Στην παράγραφο Γ.2.1 παρουσιάζονται τα βασικά στάδια της ανάλυσης ενός διαγωνίσματος.

Στις παραγράφους Γ.2.2 ως Γ.2.6 αναλύονται τα βασικότερα θέματα με τα οποία ασχολείται ο σχεδιαστής ενός C.A.T και συγκεκριμένα τα κριτήρια επιλογής αντικειμένων , μεθόδους για την αποτροπή της υπερβολικής έκθεσης των αντικειμένων , τη διαχείριση του ισοζυγίου του περιεχομένου μιας εξέτασης , τα κριτήρια τερματισμού μιας εξέτασης και τους μεθόδους για την υλοποίηση χρονικών περιορισμών σε μια εξέταση.

Γ.2.1 Ανάλυση και σχεδίαση CAT

Κατά τη διαδικασία ανάλυσης ενός διαγωνίσματος που θα χρησιμοποιηθεί για την εκτίμηση μιας ικανότητας ή κάποιου χαρακτηριστικού γνωρίσματος πρέπει να γίνουν απαραίτητα τα παρακάτω βήματα.

Αρχικά είναι σημαντικό να καθοριστεί πριν ακόμη γραφούν τα αντικείμενα , επακριβώς η ικανότητα ή το χαρακτηριστικό γνώρισμα του ατόμου που πρέπει να εκτιμάει το διαγώνισμα. Επίσης είναι σημαντικό να γίνουν αρκετά δοκιμαστικά διαγωνίσματα χρησιμοποιώντας τα αντικείμενα που υπάρχουν αρχικά στην τράπεζα αντικειμένων , ώστε να απορριφθούν αντικείμενα που δεν είναι διατυπωμένα σωστά ή έχουν κάποια ανεπιθύμητα χαρακτηριστικά . Π.χ. τα ασαφή αντικείμενα ή αντικείμενα με αρνητική διακριτική ικανότητα.

Έπειτα πρέπει να επιλεγεί ένα μοντέλο απόκρισης , το οποίο να εκφράζει τα αποτελέσματα των πειραματικών διαγωνισμάτων. Το μοντέλο αυτό χρησιμοποιείται ακολουθώντας μια μέθοδο εκτίμησης παραμέτρων για τον καθορισμό των παραμέτρων των αντικειμένων. Όταν γίνουν τα παραπάνω έχουμε στη διάθεσή μας μια τράπεζα προ-βαθμονομιμένων αντικειμένων τα οποία μπορούμε να χρησιμοποιήσουμε για τη διεξαγωγή επόμενων διαγωνισμάτων.

Εφόσον υπάρχει μια τράπεζα αντικειμένων τα οποία έχουν βαθμονομηθεί σε μια κοινή κλίμακα μπορούν να κατασκευαστούν διαγωνίσματα χρησιμοποιώντας κάποια από αυτά τα αντικείμενα. Τα διαγωνίσματα αυτά θα μετρούν την ικανότητα αυτή.

Μπορούν επίσης να χρησιμοποιηθούν διαφορετικά διαγωνίσματα ανάλογα με τις απαιτήσεις για την εκτίμηση. Για παράδειγμα αν υπάρχει η απαίτηση για μεγάλη ακρίβεια μπορούν να χρησιμοποιηθούν διαγωνίσματα πολλών αντικειμένων. Αν υπάρχει απαίτηση για ασφάλεια των αποτελεσμάτων μπορούν να χρησιμοποιηθούν πολλά διαγωνίσματα σε διαφορετικές χρονικές περιόδους και όχι μόνο ένα. Γενικά δηλαδή εφόσον υπάρχει η τράπεζα αντικειμένων μπορούν να κατασκευασθούν ειδικά τροποποιημένα διαγωνίσματα ώστε να πληρούν τις εκάστοτε απαιτήσεις για την εκτίμηση. Σε τέτοιες περιπτώσεις δεν θα χρησιμοποιηθεί όλη η τράπεζα αντικειμένων αλλά ένα μέρος της , το οποίο θα αποτελείται από τα αντικείμενα που έχουν τα απαιτούμενα τεχνικά χαρακτηριστικά. Με τον ίδιο τρόπο μπορούν να κατασκευαστούν διαγωνίσματα για την εκτίμηση μιας ικανότητας , η οποία είναι υποσύνολο της ικανότητας σύμφωνα με την οποία έχει βαθμονομηθεί η τράπεζα αντικειμένων, με την προϋπόθεση ότι το υποσύνολο των αντικειμένων της τράπεζας που θα επιλεγεί είναι καλά κατανεμημένο και το μέγεθός του επαρκεί για τις ανάγκες ενός διαγωνίσματος.

Έχοντας μια τράπεζα βαθμονομημένων αντικειμένων μπορεί να υπολογιστεί η χαρακτηριστική καμπύλη και η καμπύλη πληροφορίας ενός διαγωνίσματος πριν τη διεξαγωγή του. Με αυτό τον τρόπο μπορεί κάποιος να ξέρει από πριν πόσο «δύσκολο» είναι το διαγώνισμα αυτό και πόση πληροφορία παρέχει για τη ζητούμενη ικανότητα , δηλαδή με πόση ακρίβεια μπορεί να εκτιμηθεί η ικανότητα αυτή.

Η περιοχή ικανοτήτων στην οποία λειτουργεί σωστά ένα CAT εξαρτάται σε μεγάλο βαθμό από την περιοχή στην οποία είναι κατανεμημένα με βάση τη δυσκολία τους τα αντικείμενα που χρησιμοποιούνται σε αυτό.

Με βάση τα παραπάνω η οι κύριες κατηγορίες διαγωνισμάτων που μπορεί κανείς να κατασκευάσει είναι

- Διαγωνίσματα επιλογής
  Χρησιμοποιώντας αντικείμενα με δυσκολίες σε μια συγκεκριμένη περιοχή του άξονα μπορεί να εστιαστεί η εκτίμηση της ικανότητας σε αυτή την περιοχή. Με αυτό τον τρόπο μπορεί να επιτευχθεί σημαντική αύξηση στην ακρίβεια των αποτελεσμάτων για αυτή την περιοχή. Τέτοια διαγωνίσματα έχουν συνήθως σκοπό να διαχωρίσουν του εξεταζόμενους με ικανότητα κάτω από ένα επίπεδο , για τους οποίους ουσιαστικά δεν δίνουν εκτίμηση ικανότητας γιατί ενδεχομένως αυτή να μην ενδιαφέρει , και ταυτόχρονα να εκτιμήσουν με μεγάλη ακρίβεια την ικανότητα των υπόλοιπων εξεταζομένων. Απευθύνονται δηλαδή σε εξεταζόμενους με ικανότητες σε πολύ μικρό εύρος της κλίμακας ικανοτήτων. Για παράδειγμα διαγωνίσματα για την ανάθεση κάποιας θέσης εργασίας στον καλύτερο εξεταζόμενο που πληρεί αυστηρά συγκεκριμένες απαιτήσεις σχετικά με τη ζητούμενη ικανότητα ή η βράβευση του καλύτερου εξεταζόμενου με κάποια υποτροφία.

- Διαγωνίσματα για εξεταζόμενους με ικανότητες που εκτείνονται σε μεγάλο εύρος της κλίμακας ικανοτήτων. Για παράδειγμα τα διαγωνίσματα στα σχολεία. Τα διαγωνίσματα αυτά έχουν τη δυνατότητα να εκτιμούν την ικανότητα εξεταζομένων των οποίων οι ικανότητες εκτείνονται σε μεγαλύτερη περιοχή της κλίμακας , το σφάλμα όμως της εκτίμησης είναι σημαντικά μεγαλύτερο σε σχέση με τα διαγωνίσματα επιλογής.

Τα βήματα που εκτελούνται σε ένα CAT είναι τα παρακάτω:

1. Αρχικοποιείται η ικανότητα του καθενός από τους εξεταζόμενους σε κάποια τιμή. Η τιμή αυτή μπορεί να είναι σταθερή ή να δίνεται σύμφωνα με κάποια προκαθορισμένη κατανομή.
2. Ελέγχεται το κριτήριο τερματισμού του διαγωνίσματος.
3. Βρίσκεται χρησιμοποιώντας κάποιο κριτήριο επιλογής το επόμενο αντικείμενο που θα κληθεί να απαντήσει ο εξεταζόμενος.
4. Υπολογίζεται μια εκτίμηση της ικανότητας του εξεταζόμενου χρησιμοποιώντας κάποιον από τους αλγορίθμους εκτίμησης με βάση τις προηγούμενες απαντήσεις που έχει δώσει αυτός.
5. Επαναλαμβάνονται τα βήματα 2 ως 4 μέχρι να τελειώσει το διαγώνισμα.

Η εκτίμηση της ικανότητας του εξεταζόμενου στο βήμα 4 γίνεται με κάποια από τις μεθόδους της IRT που έχουν αναφερθεί στην παράγραφο Β.3

Γ.2.2 Κριτήρια επιλογής αντικειμένου

Υπάρχουν πολλές προσεγγίσεις για αυτό το κριτήριο με το οποίο γίνεται η επιλογή των αντικειμένων σε ένα CAT. Ένας αλγόριθμος που χρησιμοποιείται συχνά είναι αυτός που

βασίζεται στο κριτήριο της μέγιστης πληροφορίας Fisher. Σύμφωνα με αυτόν μετά από κάθε απάντηση του εξεταζόμενου εκτιμάται η ικανότητά του και χρησιμοποιώντας αυτή την εκτίμηση υπολογίζεται η πληροφορία Fisher που περιέχει κάθε αντικείμενο στην τράπεζα του CAT. Το επόμενο αντικείμενο του διαγωνίσματος επιλέγεται να είναι αυτό με τη μεγαλύτερη πληροφορία. Ουσιαστικά ο εξεταζόμενος καλείται κάθε φορά να απαντήσει στο αντικείμενο το οποίο εκτιμάται ότι θα κάνει το σφάλμα εκτίμησης το ελάχιστο δυνατό , καθώς το σφάλμα είναι αντιστρόφως ανάλογο με την πληροφορία. Αυτό που προσπαθεί να πετύχει αυτός ο αλγόριθμος είναι να μεγιστοποιήσει την πληροφορία ενός διαγωνίσματος , χρησιμοποιώντας την αθροιστική ιδιότητα της πληροφορίας και υποθέτοντας ότι το κάθε αντικείμενο είναι ανεξάρτητο.

Μια διαφορετική προσέγγιση που χρησιμοποιείται συχνά είναι η χρήση μοντέλων Bayes στο κριτήριο επιλογής. Ένας τέτοιος αλγόριθμος χρησιμοποιεί κάποιες κατανομές οι οποίες καθορίζονται εκ των προτέρων. Μερικά κριτήρια Bayes για την επιλογή αντικειμένων προτείνονται από τον Van der Linden (1996).

Γ.2.2.1 Εναλλακτικά κριτήρια επιλογής αντικειμένου.

Το σφάλμα της εκτίμησης του θ στα πρώτα στάδια ενός CAT είναι πολύ μεγαλύτερο από ότι στα τελευταία , επειδή στα πρώτα στάδια η εκτίμηση γίνεται με πολύ λιγότερα αντικείμενα. Ειδικά για την επιλογή του πρώτου αντικειμένου η τιμή του θ αρχικοποιείται συνήθως σε μια κοινή τιμή για όλους τους εξεταζόμενους , η οποία συνήθως βρίσκεται στο μέσο της κλίμακας των ικανοτήτων. Επειδή η επιλογή των αντικειμένων βασίζεται στην εκτίμηση του θ , όσο πιο μικρό είναι το σφάλμα στην εκτίμηση αυτή τόσο πιο σωστή θα είναι και η επιλογή του επόμενου αντικειμένου. Στα πρώτα στάδια λοιπόν ενός CAT , όπου υπάρχει μεγάλο σφάλμα εκτίμησης , τα αντικείμενα που επιλέγονται δεν είναι τόσο κατάλληλα όσο τα τελευταία (Chen , Ankenmann & Chang 2000).

Από τα παραπάνω είναι φανερό ότι το κριτήριο επιλογής αντικειμένου με βάση τη μέγιστη πληροφορία μάλλον δεν λειτουργεί τόσο αποδοτικά για τα πρώτα αντικείμενα , επειδή η πληροφορία που περιέχει ένα αντικείμενο βασίζεται στην εκτίμηση της ικανότητας. Αν δηλαδή υπάρχει μεγάλο σφάλμα εκτίμησης είναι πολύ πιθανό το αντικείμενο που θα επιλεγεί να μην περιέχει πραγματικά τη μέγιστη πληροφορία.

Μια τροποποίηση του κριτηρίου επιλογής λοιπόν για τα πρώτα βήματα ενός CAT θα μπορούσε να προσφέρει μεγαλύτερη ακρίβεια όταν υπάρχουν λίγα αντικείμενα. Από τη δεκαετία του 90 έχουν προταθεί αρκετοί αλγόριθμοι επιλογής αντικειμένου που τροποποιούν το κριτήριο μέγιστης πληροφορίας , με σκοπό την ταχύτερη σύγκλιση ενός CAT. Για παράδειγμα ο Van der Linden (1995) προτείνει μια παραλλαγή του κριτηρίου μέγιστης πληροφορίας χρησιμοποιώντας μια προσέγγιση μοντέλων Bayes και δείχνει ότι τροποποιημένη μέθοδος επιλογής έχει θεωρητικά καλύτερη απόδοση από το κριτήριο μέγιστης πληροφορίας.

Συγκεκριμένα θεωρεί μια εκ των προτέρων κατανομή που εκτείνεται σε όλο το εύρος της κλίμακας του θ, την οποία χρησιμοποιεί σα βάρος στη συνάρτηση πληροφορίας Fisher στον

υπολογισμό των πληροφοριών του κάθε αντικειμένου. Η κατανομή έχει το μέγιστο της στο μέσο της κλίμακας των ικανοτήτων αρχικά. Με αυτό τον τρόπο το αντικείμενο με τη μεγαλύτερη πληροφορία έχει τη μεγαλύτερη πιθανότητα , χωρίς όμως να αποκλείονται και άλλα αντικείμενα. Χρησιμοποιώντας μια κατανομή Gauss , όσο μεγαλύτερη είναι δηλαδή η πληροφορία ενός αντικειμένου , τόσο μεγαλύτερη είναι και η πιθανότητα να επιλεγεί αυτό.

Οι Shu-Ying Chen ,Robert D. Ankenmann, Hua-Hua Chang (2000) συγκρίνοντας την παραπάνω με άλλες 3 εναλλακτικές μεθόδους και με τη μέθοδο μέγιστης πληροφορίας βρίσκουν ότι πρακτικά οι τέσσερις εναλλακτικές μέθοδοι δίνουν καλύτερα αποτελέσματα μόνο για πολύ μεγάλες αρνητικές τιμές θ σε διαγωνίσματα λίγων αντικειμένων . Για διαγωνίσματα πάνω των 10 αντικειμένων όλες οι μέθοδοι αποδίδουν παρόμοια σε όλα τα μεγέθη που ενδιαφέρουν και για όλα τα θ.

Οι Wainer , Kaplan και Lewis (1992) πρότειναν έναν αλγόριθμο επιλογής ερωτήσεων για τις προσαρμοστικές εξετάσεις έτσι ώστε να μεγιστοποιείται ο διαχωρισμός μεταξύ των εξεταζομένων , χωρίς τη χρήση της IRT. Ο αλγόριθμος αυτός όμως βρέθηκε ότι υστερεί σε σχέση με τον αλγόριθμο μέγιστης πληροφορίας με βάση την IRT (Schnipke & Green 1995).
Συγκεκριμένα οι Schnipke & Green δοκιμάζοντας και τους δύο αλγορίθμους σε ένα κανονικά κατανεμημένο δείγμα 1000 εξεταζόμενων καταλήγουν στο συμπέρασμα ότι τα διαγωνίσματα που χρησιμοποιούν τον αλγόριθμο μέγιστης πληροφορίας έδωσαν την περισσότερη πληροφορία για μεγαλύτερο εύρος τιμών ικανότητας και σε γενικές γραμμές διαχώρισαν τους εξεταζόμενους καλύτερα από τα διαγωνίσματα που χρησιμοποιούσαν τον άλλο αλγόριθμο. Το συμπέρασμα από την παραπάνω έρευνα είναι ότι οι προσαρμοστικές εξετάσεις που βασίζονται στο κριτήριο της μέγιστης πληροφορίας υπερέχουν ξεκάθαρα.

Γ.2.3 Έκθεση αντικειμένων (Item Exposure)

Ένα από τα σημαντικότερα προβλήματα των CAT είναι η υπερβολική χρησιμοποίηση (έκθεση) κάποιων αντικειμένων. Ανάλογα με το κριτήριο επιλογής αντικειμένου που χρησιμοποιείται , ορισμένα αντικείμενα τείνουν να επιλέγονται συχνότερα από άλλα. Για παράδειγμα σε ένα CAT που υποθέτει αρχικά θ=0 για όλους τους εξεταζόμενους και χρησιμοποιεί το κριτήριο μέγιστης πληροφορίας Fisher η πρώτη ερώτηση θα είναι κοινή για όλους τους εξεταζόμενους. Αν η τράπεζα αντικειμένων δεν αλλάξει σε επόμενες εξετάσεις η πρώτη ερώτηση θα είναι η ίδια. Το πρόβλημα έγκειται στην έγκυρη αξιολόγηση των εξεταζομένων και γενικά την ασφάλεια της εξέτασης. Όταν ορισμένα αντικείμενα χρησιμοποιούνται πολύ συχνά είναι πιθανό ένας εξεταζόμενος να τα γνωρίζει πριν την εξέταση , είτε διότι έχει εξεταστεί ξανά είτε γιατί έχει συλλέξει πληροφορίες από άτομα που έχουν ήδη εξεταστεί.

Επιπλέον αν ορισμένα αντικείμενα δεν επιλέγονται σχεδόν ποτέ , τότε έχει δαπανηθεί χρόνος από τους αναλυτές και χρήμα από τους εργοδότες χωρίς σκοπό. Αν δεν υπάρχει κάποιος αποδοτικός τρόπος χρησιμοποίησης της τράπεζας , τότε για λόγους εγκυρότητας και ασφάλειας αυτή πρέπει να ανανεώνεται σε τακτά χρονικά διαστήματα , απαιτώντας επιπλέον χρόνο ανάλυσης και σχεδίασης και προφανώς χρήμα. Όταν χρησιμοποιείται το

κριτήριο της μέγιστης πληροφορίας τα αντικείμενα που έχουν μεγάλες τιμές παραμέτρου α τείνουν να χρησιμοποιούνται περισσότερο από τα άλλα.

Οι Georgiadou , Triantafillou και Economides (2007) κάνουν μια επισκόπηση των μεθόδων για τον έλεγχο της έκθεσης αντικειμένων καθώς και των μελετών που τις συγκρίνουν ως προς την απόδοση. Σύμφωνα με την παραπάνω μελέτη οι μέθοδοι αυτές μπορούν να χωριστούν στις ακόλουθες κατηγορίες.

- **Randomization** οι οποίοι βασίζονται στην τυχαία επιλογή ενός από ένα σύνολο αντικειμένων των οποίων η πληροφορία είναι κοντά στη μέγιστη)
- **Conditional Selection** οι περισσότεροι από τους οποίους χρησιμοποιούν κάποια παράμετρο έκθεσης η οποία συνυπολογίζεται κατά την επιλογή αντικειμένου
- **Stratified Strategies** οι οποίοι κατανέμουν τα αντικείμενα σε στρώματα και χρησιμοποιούν το κάθε αντικείμενο στο στάδιο του διαγωνίσματος που αυτό είναι πιο χρήσιμο.
- **Combined Strategies** οι οποίοι συνδυάζουν διάφορες από τις παραπάνω μεθόδους
- **Multiple Stage Adaptive Test Designs** οι οποίοι προσπαθούν να ελέγξουν την έκθεση των αντικειμένων εκ των προτέρων και όχι να τροποποιήσουν το κριτήριο επιλογής αντικειμένου.

Ένα στατιστικό μέγεθος που βοηθά στη μέτρηση της έκθεσης των αντικειμένων είναι ο ρυθμός έκθεσης αντικειμένου (item exposure rate) ο οποίος είναι ο λόγος του αριθμού των εξεταζόμενων που έχουν απαντήσει στο αντικείμενο αυτό κατά τη διάρκεια μιας εξέτασης προς το συνολικό αριθμό των εξεταζόμενων. Το μέγεθος αυτό μας δείχνει κατά πόσο χρησιμοποιείται ένα αντικείμενο. Όταν υπάρχουν αντικείμενα με μικρό ρυθμό έκθεσης (Under-exposed items), τότε σημαίνει ότι η τράπεζα αντικειμένων δε χρησιμοποιείται αποτελεσματικά. Το ίδιο συμβαίνει βέβαια και αν υπάρχουν αντικείμενα με μεγάλο ρυθμό έκθεσης (over exposed items), το οποίο σημαίνει ότι χρησιμοποιούνται πολύ συχνά.

Γ.2.4 Διαχείριση Ισοζυγίου του περιεχομένου της εξέτασης(Content Balancing)

Ένα άλλο πρόβλημα που πρέπει να λαμβάνεται υπ όψη στην επιλογή αντικειμένων είναι η ισοστάθμιση του περιεχομένου (content balancing) της εξεταστέας ύλης σε ένα διαγώνισμα. Αν για παράδειγμα σκοπός μιας εξέτασης είναι η εκτίμηση της ικανότητας στα μαθηματικά και η εξεταστέα ύλη περιλαμβάνει ολοκλήρωση , διαφόριση και ακολουθίες για να εκτιμηθεί η ικανότητα ενός εξεταζομένου στα μαθηματικά θα πρέπει να υπάρχουν αντικείμενα από όλες τις θεματικές ενότητες. Στη γενική περίπτωση , όταν δε δίνεται ιδιαίτερο βάρος σε κάποια από τις θεματικές ενότητες , τα αντικείμενα πρέπει να είναι κατανεμημένα ομοιόμορφα σε αυτές. Για να επιτευχθεί αυτό πρέπει να λαμβάνονται υπ όψη και οι θεματικές ενότητες στις οποίες ανήκουν τα αντικείμενα από το κριτήριο επιλογής αντικειμένων. Είναι κατανοητό ότι βάζοντας ένα τέτοιο περιορισμό στην επιλογή αντικειμένου θα υπάρχουν μεγαλύτερα σφάλματα στην εκτίμηση της ικανότητας. Για να μειωθεί η αύξηση του σφάλματος θα πρέπει κατά τη δημιουργία της τράπεζας αντικειμένων να ληφθούν υπ όψη και οι θεματικές ενότητες. Θα πρέπει δηλαδή όχι μόνο οι παράμετροι των αντικειμένων να είναι σωστά κατανεμημένες ανάλογα με τις απαιτήσεις

της εξέτασης , αλλά και τα αντικείμενα να είναι ομοιόμορφα κατανεμημένα στις επιμέρους θεματικές ενότητες ανάλογα με τις παραμέτρους τους. Στο παραπάνω παράδειγμα δε θα ήταν σωστό να υπάρχουν μόνο εύκολες ερωτήσεις ολοκλήρωσης και μόνο δύσκολες διαφόρισης.

Η μέθοδος που χρησιμοποιείται συχνότερα για την ισοστάθμιση του περιεχομένου είναι η μέθοδος constrained CAT (CCAT) που προτάθηκε από τους Kingsbury , Zara (1989). Σύμφωνα με αυτή τη μέθοδο ορίζεται για την κάθε θεματική ενότητα ένας αριθμός που αντιστοιχεί στο μέγιστο αριθμό ζητούμενων αντικειμένων από τη συγκεκριμένη ενότητα στο διαγώνισμα και εκφράζεται σαν ποσοστό των συνολικών αντικειμένων του διαγωνίσματος. Μετά την επιλογή κάποιου αντικειμένου μειώνεται ο αριθμός αυτός κατά μια μονάδα. Στο κριτήριο επιλογής αντικειμένου , το αντικείμενο επιλέγεται από την ή τις ενότητες με το μέγιστο αριθμό ζητούμενων αντικειμένων. Πρακτικά δηλαδή αν απαιτείται ίδιος αριθμός ερωτήσεων από όλες τις θεματικές ενότητες , χρησιμοποιώντας αυτό τον αλγόριθμο αν έχουμε τρείς ενότητες τα πρώτα τρία αντικείμενα θα ανήκουν το καθένα σε μια από τις τρεις όπως και τα επόμενα τρία κ.ο.κ.

Οι Chen,Ankenmann(2004) πρότειναν μια λιγότερο προβλέψιμη λύση χρησιμοποιώντας ένα πολυωνυμικό μοντέλο (modified multinomial model ή MMM). Σύμφωνα με αυτή τη μέθοδο δημιουργείται αρχικά μια κοινή κατανομή για όλες τις θεματικές ενότητες με βάση τα ζητούμενα ποσοστά αντικειμένων από την κάθε ενότητα. Επιλέγεται ένας τυχαίος αριθμός από μια κανονική κατανομή U(0,1). Ο αριθμός αυτός μέσω της κοινής κατανομής καθορίζει την θεματική ενότητα στην οποία θα ανήκει το αντικείμενο που θα επιλεγεί από τον αλγόριθμο επιλογής αντικειμένου. Όταν για μια θεματική ενότητα το ποσοστό των αντικειμένων της σε σχέση με τα συνολικά είναι ίσο με το ζητούμενο δημιουργείται μια νέα πολυωνυμική κατανομή για τις θεματικές ενότητες. Με τον τρόπο αυτό δείχνουν ότι η σειρά των θεματικών ενοτήτων με την οποία επιλέγονται τα αντικείμενα είναι τυχαία και μη προβλέψιμη.

Μια άλλη προσέγγιση για τη βελτιστοποίηση του CCAT είναι η Modified Constrained CAT (MCCAT) Που προτάθηκε από τους Leung et al. (2003). Σύμφωνα με αυτή την τροποποίηση του CCAT στο κριτήριο επιλογής αντικειμένου , το αντικείμενο επιλέγεται από την ή τις ενότητες για τις οποίες ο ζητούμενος αριθμός αντικειμένων δεν είναι μηδέν. Πρακτικά δηλαδή επιλέγονται τα βέλτιστα αντικείμενα μέχρι για κάποια ενότητα να συμπληρωθεί το απαιτούμενο ποσοστό αντικειμένων. Η επιλογή των αντικειμένων στη συνέχεια γίνεται μόνο από τις άλλες ενότητες.

Οι Leung, Chang, & Hau, 2003 σύγκριναν τις τρεις μεθόδους χρησιμοποιώντας το κριτήριο επιλογής μέγιστης πληροφορίας και βρήκαν ότι οι τρείς μέθοδοι
Αποδίδουν παρόμοια όσο αφορά την ακρίβεια της εκτίμησης , τα σφάλματα εκτίμησης και την απόκλιση (bias) όσο και τη συσχέτιση των εκτιμούμενων τιμών θ με τις πραγματικές.
Όσο αφορά την εκμετάλλευση της τράπεζας των αντικειμένων υπάρχουν πολύ μικρές διαφορές οι οποίες μειώνονται όσο αυξάνεται ο αριθμός των αντικειμένων σε ένα διαγώνισμα. Η MMM βρέθηκε ότι δίνει τα λιγότερα over-exposed αντικείμενα και η CCAT τα περισσότερα. Σε γενικές γραμμές όμως οι τρεις μέθοδοι αποδίδουν παρόμοια.

Τα τελευταία χρόνια γίνονται έρευνες για την ανάπτυξη ευέλικτων μεθόδων ισοστάθμισης περιεχομένου (flexible content balancing) , πολλές από τις οποίες στηρίζονται στις τρεις παραπάνω μεθόδους με σκοπό τη βελτίωση της απόδοσής τους ως προς τη χρήση της τράπεζας αντικειμένων.
Οι τρεις παραπάνω αλγόριθμοι χαρακτηρίζονται στατικοί γιατί το ποσοστό των αντικειμένων από κάθε ενότητα είναι προκαθορισμένο και σταθερό. Οι Flexible αλγόριθμοι αντίθετα ορίζουν άνω και κάτω όρια για το ποσοστό των αντικειμένων από κάθε κεφάλαιο. Τέσσερις τέτοιες μέθοδοι για παράδειγμα προτείνονται από τους Cheng & Chang (2007).

Στην ιδανική περίπτωση θα ήθελε κάποιος ένα CAT που να επιτυγχάνει ταυτόχρονα μεγάλη ακρίβεια στην εκτίμηση της ικανότητας , σωστή κατανομή των αντικειμένων στις θεματικές ενότητες της εξεταστέας ύλης και όσο το δυνατό καλύτερη χρησιμοποίηση της τράπεζας αντικειμένων ώστε να μην υπάρχουν αντικείμενα που χρησιμοποιούνται πολύ συχνά. Οι τρεις αυτοί στόχοι όμως είναι αντικρουόμενοι μεταξύ τους. Θέτοντας για παράδειγμα κάποιους περιορισμούς στην επιλογή αντικειμένου έτσι ώστε να υπάρχει σωστότερη κατανομή των αντικειμένων στις θεματικές ενότητες , αυξάνεται ταυτόχρονα και το σφάλμα εκτίμησης και η πιθανότητα για υπερβολική χρησιμοποίηση κάποιων αντικειμένων.

Γ.2.5 Κριτήρια τερματισμού του διαγωνίσματος

Σκοπός ενός CAT είναι η ακριβής εκτίμηση της ικανότητας των εξεταζόμενων. Έτσι ένα από τα κριτήρια που μπορεί να χρησιμοποιηθεί είναι ένα όριο στο σφάλμα εκτίμησης το οποίο θεωρεί ο εξεταστής ικανοποιητικό. Όταν μετά από κάποια απάντηση του εξεταζομένου το σφάλμα υπολογιστεί μικρότερο από αυτή την τιμή η εξέταση μπορεί να τελειώσει καθώς το σφάλμα εκτίμησης στους περισσότερους αλγόριθμους μειώνεται καθώς αυξάνεται ο αριθμός των αντικειμένων

Ένα άλλο κριτήριο τερματισμού είναι ο χρόνος εξέτασης. Για πρακτικούς λόγους η κάθε εξέταση μπορεί να διαρκεί ορισμένο χρονικό διάστημα , έτσι είναι απαραίτητο να τεθεί κάποιο χρονικό όριο για τη διάρκειά της. Για την απόφαση αυτού του ορίου πρέπει να ληφθούν υπ όψη οι χρόνοι που απαιτούνται για την απάντηση των αντικειμένων , έτσι ο μέσος εξεταζόμενος , αν όχι όλοι οι εξεταζόμενοι , να έχει τον απαραίτητο χρόνο να απαντήσει σε ικανοποιητικό αριθμό αντικειμένων για να εκτιμηθεί η ικανότητά του με ακρίβεια.

Ένα τρίτο κριτήριο είναι ο αριθμός αντικειμένων σε ένα διαγώνισμα. Πρέπει να υπάρχει ένας περιορισμός μέγιστου αριθμού αντικειμένων για ένα διαγώνισμα έτσι ώστε αυτό τερματίζεται στην περίπτωση που δεν χρησιμοποιείται κάποιο κριτήριο χρόνου.

Συνήθως χρησιμοποιείται κάποιος συνδυασμός των κριτηρίων αυτών για την καλύτερη διεξαγωγή των εξετάσεων και ανάλογα με τις εκάστοτε απαιτήσεις για ακρίβεια εκτίμησης.

Γ.2.6 Περιορισμοί χρόνου.

Οι πρακτικοί περιορισμοί κάθε διαγωνίσματος όπως ο χρόνος και ο μέγιστος αριθμός των αντικειμένων εισάγουν επιπλέον προβλήματα στην αξιολόγηση της επίδοσης τόσο στην κλασική μέθοδο αξιολόγησης όσο και στα CAT.

Στον τομέα της εκπαίδευσης η ταχύτητα απόκρισης δεν είναι κάτι που ενδιαφέρει άμεσα όπως σε άλλους τομείς. Για παράδειγμα στην αξιολόγηση της απόδοσης ενός εργαζομένου η ταχύτητα παίζει σημαντικό ρόλο. Σε κάθε περίπτωση όμως η εκτίμηση της ικανότητας ενός εξεταζόμενου εξαρτάται και από την ταχύτητα απόκρισής του. Ένα από τα σημαντικότερα προβλήματα που εισάγει ο περιορισμός του χρόνου διεξαγωγής μιας εξέτασης είναι λοιπόν η ακρίβεια της εκτίμησης της ικανότητας στην περίπτωση που δεν υπάρχουν αρκετά δεδομένα. Για παράδειγμα η εκτίμηση της ικανότητας ενός εξεταζόμενου που έχει απαντήσει μόνο σε δύο δύσκολα αντικείμενα από τα 20 ενός διαγωνίσματος.

Στην κλασική θεωρία συνήθως αντικείμενα που δεν έχουν απαντηθεί θεωρείται ότι έχουν απαντηθεί λανθασμένα. Δεν υπάρχει άλλωστε κάποιος τρόπος να γνωρίζει ο βαθμολογητής εάν ο εξεταζόμενος δεν πρόλαβε να απαντήσει στα αντικείμενα αυτά ή δεν γνώριζε να απαντήσει. Αντίθετα όταν η εξέταση γίνεται μέσω υπολογιστή είναι γνωστό ποια αντικείμενα πρόλαβε και ποια όχι να δει ο εξεταζόμενος. Υπάρχει δηλαδή η δυνατότητα τα αντικείμενα δεν έχουν απαντηθεί να θεωρηθούν δεδομένα εισόδου που λείπουν στην εκτίμηση της ικανότητας , και η εκτίμηση αυτή να γίνει σύμφωνα με τα δεδομένα που υπάρχουν. Στην περίπτωση αυτή και όταν χρησιμοποιείται CAT θα υπάρχει μεγάλο σφάλμα εκτίμησης. Όσο λιγότερα είναι τα αντικείμενα που έχει απαντήσει ένας εξεταζόμενος τόσο μεγαλύτερο θα είναι το σφάλμα. Κατά τη σχεδίαση και την ανάλυση ενός CAT πρέπει να λαμβάνεται υπ’ όψη και ο χρόνος που απαιτείται για την απάντηση των ερωτήσεων και να ορίζεται ανάλογα η χρονική διάρκεια της εξέτασης. Παρ’ όλα αυτά υπάρχει πάντα το ενδεχόμενο να υπάρχουν ελειπή δεδομένα και το σφάλμα εκτίμησης να είναι μεγάλο. Σε τέτοιες περιπτώσεις πρέπει να χρησιμοποιείται ένας διαφορετικός τρόπος για την εκτίμηση ή να μη δίνεται εκτίμηση λόγω ελειπών δεδομένων.

Μελέτες που έχουν γίνει έχουν δείξει ότι δεν υπάρχει μεγάλη συσχέτιση μεταξύ της ικανότητας του εξεταζόμενου , της δυσκολίας ενός αντικειμένου και του χρόνου που απαιτείται για να απαντήσει ο εξεταζόμενος στο αντικείμενο αυτό. Δεν υπάρχει κάποιο μοντέλο δηλαδή , που να προσομοιώνει τη συμπεριφορά των εξεταζόμενων ως προς το χρόνο σε σχέση με τη δυσκολία των αντικειμένων , ούτε σε σχέση με την ικανότητά τους. Πρακτικά οι χρόνοι απόκρισης σε ένα αντικείμενο για δύο εξεταζόμενους της ίδιας ικανότητας μπορεί να έχουν μεγάλη διαφορά. Επιπλέον κάποιος εξεταζόμενος μπορεί να χρειαστεί περισσότερο χρόνο για να απαντήσει σε ένα ευκολότερο αντικείμενο από ότι σε ένα πιο δύσκολο , χωρίς αυτός ο χρόνος να εξαρτάται αποκλειστικά στη δυσκολία του αντικειμένου. Ο χρόνος απόκρισης ενός εξεταζόμενου στο ίδιο αντικείμενο εξαρτάται και από άλλους παράγοντες που δεν μπορούν να παραμετροποιηθούν όπως το πόσο συγκεντρωμένος είναι , πόσο καλά γνωρίζει το αντικείμενο της εξέτασης , πόσο εύκολα μπορεί να κατανοεί ένα κείμενο , πόσο καλά γνωρίζει τη γλώσσα στην οποία είναι διατυπωμένες οι ερωτήσεις , ψυχολογικούς λόγους , κούραση κτλ. Δεν υπάρχει λοιπόν

κάποιος τρόπος να προβλεφθεί με μεγάλη ακρίβεια ο χρόνος απόκρισης ενός εξεταζόμενου από τη δυσκολία του αντικειμένου και μια εκτίμηση της ικανότητάς του. Αν μπορούσε δυναμικά να προβλέπεται ο χρόνος που θα κάνει ένας εξεταζόμενος να απαντήσει σε ένα σύνολο ερωτήσεων θα μπορούσε να τροποποιηθεί το κριτήριο επιλογής αντικειμένου σε ένα CAT ώστε να υπάρχει τελικά περισσότερη πληροφορία σε περιπτώσεις ελλειπών δεδομένων και συνεπώς μικρότερο σφάλμα εκτίμησης. Επειδή όμως σε ένα CAT η ικανότητα του εξεταζόμενου εκτιμάται σε κάθε βήμα , και το σφάλμα αυτής της εκτίμησης είναι αρκετά μεγάλο ειδικά στα πρώτα βήματα της εξέτασης, το να γίνει μια σωστή πρόβλεψη όσο αφορά το χρόνο που θα κάνει ο εξεταζόμενος να απαντήσει στις απαιτούμενες ερωτήσεις είναι πολύ δύσκολο.

Γ.3 Σύγκριση C.A.T με την κλασική μέθοδο εξέτασης

Το σημαντικότερο πλεονέκτημα ενός C.A.T έναντι της κλασικής μεθόδου εξέτασης είναι η μεγάλη ακρίβεια εκτίμησης. Σύμφωνα με την ανάλυση που έχει γίνει στις προηγούμενες παραγράφους για να επιτευχθεί η ακρίβεια εκτίμησης που δίνει ένα C.A.T με την κλασική μέθοδο απαιτείται μεγάλος αριθμός αντικειμένων , δηλαδή διαγωνίσματα περισσότερων ερωτήσεων και συνεπώς μεγαλύτερης χρονικής διάρκειας , τα οποία πολλές φορές δεν είναι εφικτό να πραγματοποιηθούν. Στην κλασική μέθοδο ο εξεταζόμενος πρέπει να απαντήσει σε σημαντικό αριθμό ερωτήσεων για να επιτευχθεί ικανοποιητική ακρίβεια εκτίμησης , καθώς αυτές είναι κοινές για όλους , απαντώντας έτσι σε πολλές ερωτήσεις που δίνουν ελάχιστη ως καθόλου πληροφορία για την ικανότητά του. Αντίθετα στα C.A.T ο εξεταζόμενος απαντά σε όσες ερωτήσεις είναι απαραίτητες για την απαιτούμενη ακρίβεια εκτίμησης. Τα διαγωνίσματα που απαιτούν λιγότερο χρόνο εξέτασης επίσης είναι ένα μεγάλο πλεονέκτημα , ειδικά σε περιπτώσεις που η εξέταση χρησιμοποιείται για ιατρικούς σκοπούς , καθώς ο εξεταζόμενος καταπονείται λιγότερο.

Χρησιμοποιώντας C.A.T διασφαλίζεται η εγκυρότητα και η ασφάλεια της εξέτασης και των αποτελεσμάτων. Το διαγώνισμα είναι διαφορετικό για κάθε εξεταζόμενο και έτσι μπορεί να μειωθεί το ποσοστό των αντιγραφών. Πρέπει βέβαια να γίνεται σωστή χρήση της τράπεζας των αντικειμένων ώστε να μη χρησιμοποιούνται υπερβολικά κάποια από αυτά και να ανανεώνεται η τράπεζα ώστε να μην είναι γνωστές οι ερωτήσεις στους εξεταζόμενους.

Η βαθμολόγηση των αποτελεσμάτων στην κλασική μέθοδο απαιτεί σημαντικό χρόνο και κόπο από τον εξεταστή. Αντίθετα σε ένα CAT γίνεται από τον υπολογιστή και έτσι απαιτεί ελάχιστο χρόνο και κόπο από τον εξεταστή.

Το CAT δίνει στον εξεταστή τη δυνατότητα να διακρίνει πολύ εύκολα κάποια προβληματικά αντικείμενα , όπως ασαφής ερωτήσεις , ερωτήσεις που μπερδεύουν τους εξεταζόμενους κτλ. Δίνεται έτσι η δυνατότητα για ποιοτικότερη εξέταση απομακρύνοντας αρκετά από τα προβλήματα που μπορούν να προκύψουν.

Οι μέθοδοι Content Balancing και ο αποδοτικότερος χειρισμός των ερωτήσεων ενός CAT δίνουν τη δυνατότητα καλύτερης κατανομής των ερωτήσεων ενός διαγωνίσματος στην

εξεταστέα ύλη και εξέτασης μεγαλύτερου εύρους ύλης. Επιπλέον δίνεται η δυνατότητα εκτίμησης για κάθε κεφάλαιο ξεχωριστά. Έτσι ο εξεταζόμενος μπορεί να ενημερωθεί άμεσα για τη απόδοσή του σε κάθε κεφάλαιο ξεχωριστά ή για τη συνολική απόδοσή του.

Μια σημαντική διαφορά ανάμεσα στις δύο μεθόδους είναι ο τρόπος που γίνεται η εκτίμηση της ικανότητας του εξεταζόμενου. Η εκτίμηση αυτή στο CAT δεν εξαρτάται από το συγκεκριμένο διαγώνισμα στο οποίο θα κληθεί να απαντήσει ο εξεταζόμενος. Αυτό συμβαίνει γιατί σύμφωνα με την IRT η εκτίμηση της ικανότητας ενός εξεταζόμενου θα είναι η ίδια αν ο εξεταζόμενος κληθεί να απαντήσει σε πολλά τεστ. Η εκτίμηση της ικανότητας δεν εξαρτάται από τη δυσκολία του διαγωνίσματος. Αντίθετα χρησιμοποιώντας την κλασική θεωρία η εκτίμηση αυτή εξαρτάται από τη δυσκολία του διαγωνίσματος. Ο ίδιος εξεταζόμενος θα απαντήσει σε λιγότερες ερωτήσεις σωστά σε ένα δυσκολότερο διαγώνισμα από ότι σε ένα πιο εύκολο. Επειδή στην κλασική θεωρία η τελική εκτίμηση της ικανότητας προκύπτει αθροιστικά από τη βαθμολογία των επιμέρους ερωτήσεων του διαγωνίσματος η εκτιμήσεις της ικανότητας για τον ίδιο εξεταζόμενο σε ένα δυσκολότερο διαγώνισμα και σε ένα πιο εύκολο θα είναι διαφορετικές.

Και οι δύο μέθοδοι εξέτασης απαιτούν χρόνο προετοιμασίας ενός διαγωνίσματος από τον εξεταστή. Στην κλασική μέθοδο πρέπει να γράψει τις ερωτήσεις , να τις κατανέμει σωστά στην εξεταστέα ύλη , και να κάνει μια υποκειμενική εκτίμηση της δυσκολίας τους. Όλα αυτά πρέπει να γίνουν πριν κάθε διαγώνισμα. Αντίθετα στο CAT πρέπει να γίνουν οι διαδικασίες που απαιτούνται για τη δημιουργία μια βαθμονομημένης τράπεζας αντικειμένων, η οποία περιλαμβάνει τη συγγραφή των αντικειμένων τη διοργάνωση δοκιμαστικών διαγωνισμάτων και την ανάλυση των αποτελεσμάτων. Αυτή η διαδικασία όμως γίνεται μέσω υπολογιστών όπως και η εκτίμηση των παραμέτρων των αντικειμένων. Επιπλέον η ίδια τράπεζα αντικειμένων μπορεί να χρησιμοποιηθεί για τη διεξαγωγή πολλών CAT , συνεπώς η εξέταση μέσω CAT απαιτεί στη γενική περίπτωση λιγότερο χρόνο προετοιμασίας από τον εξεταστή.

Ένα από τα αρνητικά του CAT σε σχέση με την κλασική μέθοδο είναι ότι ο εξεταζόμενος δεν έχει τη δυνατότητα να διαβάσει από πριν όλες τις ερωτήσεις και να κατανέμει το χρόνο της εξέτασης όπως νομίζει καλύτερα. Επιπλέον δεν του δίνεται η δυνατότητα να αλλάξει κάποια από τις απαντήσεις που έχει δώσει σε προηγούμενες ερωτήσεις. Ένα τρίτο μειονέκτημα είναι ότι επειδή στο CAT ο έλεγχος της ορθότητας των απαντήσεων γίνεται από υπολογιστή οι απαντήσεις πρέπει να δίνονται με τη μορφή επιλογής μια ή περισσότερων εναλλακτικών απαντήσεων. Δε δίνεται η δυνατότητα δηλαδή στον εξεταζόμενο να διατυπώσει την απάντησή του όπως αυτός θέλει καθώς οι απαντήσεις είναι διατυπωμένες ήδη. Αυτό εισάγει και τον παράγοντα της τύχης που δεν είναι εύκολο να προβλεφτεί και να συνυπολογιστεί στην εκτίμηση της ικανότητας. Επιπλέον ο χρήστης μπορεί να δώσει τη σωστή απάντηση σε κάποιο αντικείμενο αποκλείοντας τις άλλες εναλλακτικές.

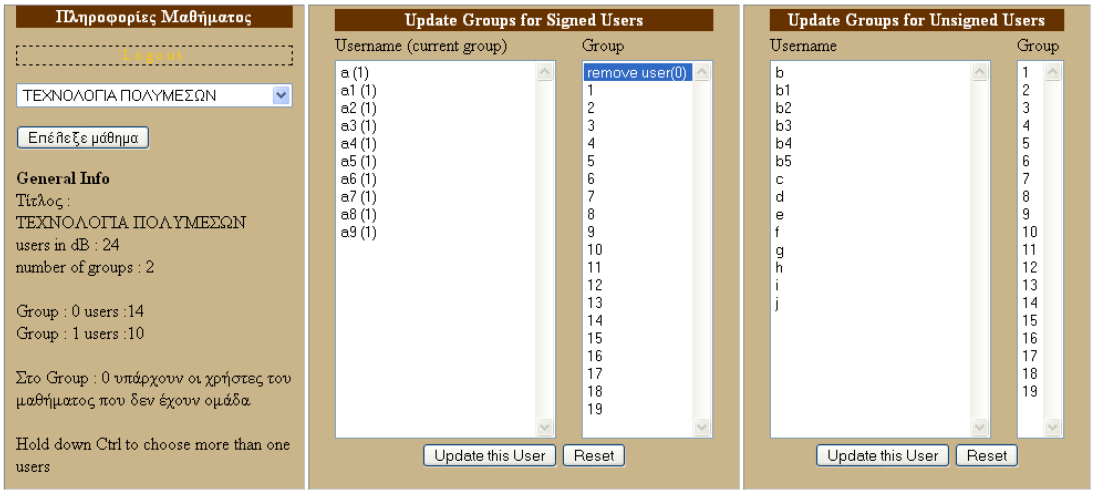
Δ. Ανάπτυξη ενός C.A.T.

Η πλατφόρμα εξέτασης έχει υλοποιηθεί σε PHP 5 ,MySQL 5 και Apache 2. Επειδή έχει χρησιμοποιηθεί τεχνολογία A.J.A.X. ο περιηγητής που θα χρησιμοποιούν οι εξεταζόμενοι πρέπει να υποστηρίζει την τεχνολογία αυτή. Περιηγητές που υποστηρίζουν κλήσεις AJAX είναι οι Firefox (όλες οι εκδόσεις), Internet Explorer (5.0 ή νεώτερος), Apple Safari (1.2 ή νεώτερος), Konqueror, Netscape (7.1 ή νεώτερος), και Opera (7.6 ή νεώτερος).

Έχουν υλοποιηθεί δύο συστήματα εξέτασης. Το ένα για την πραγματοποίηση δοκιμαστικών εξετάσεων που θα βοηθήσουν στον υπολογισμό των παραμέτρων των αντικειμένων και ένα CAT. Η πλατφόρμα admin2 βοηθάει στην οργάνωση των δοκιμαστικών εξετάσεων δίνοντας τη δυνατότητα στον εξεταστή να χωρίσει τους εξεταζόμενους σε ομάδες. Η πλατφόρμες estimate1 και estimate2 υπολογίζουν τις παραμέτρους για τα μοντέλα 1PL και 2PL αντίστοιχα. Για το CAT χρησιμοποιήθηκε το μοντέλο 2PL , αλλά μπορεί ο εξεταστής να το χρησιμοποιήσει και για τράπεζα αντικειμένων του μοντέλου 1PL , θέτοντας την παράμετρο διακριτικής ικανότητας α=1 για όλα τα αντικείμενα της τράπεζας.

Για την υλοποίηση χρησιμοποιήθηκε συνδυασμός αντικειμενοστραφούς και γραμμικού προγραμματισμού. Για τη σύνδεση με τη βάση και τις διάφορες λειτουργίες που προσφέρει αυτή χρησιμοποιήθηκε μια Wrapper κλάση , η MySQl , η οποία είναι wrapper για τη σύνδεση και τις λειτουργίες(Select , insert , delete κτλ) της MySql. Με αυτό τον τρόπο αν αλλάξει το πρόγραμμα διαχείρισης της βάσης πρέπει να αλλαχθεί μόνο αυτή η κλάση και όχι όλο το πρόγραμμα.

Στο σύστημα για την πραγματοποίηση δοκιμαστικών εξετάσεων οι εξεταζόμενοι χωρίζονται από τον εξεταστή σε ομάδες μέσω της σελίδας admin2(Σχήμα Σ.14). Σε κάθε ομάδα αντιστοιχίζεται μια ομάδα ερωτήσεων στις οποίες θα απαντήσει. Οι ερωτήσεις είναι ίδιες για τους εξεταζόμενους μιας ομάδας και εμφανίζονται κάθε εξεταζόμενο με τυχαία σειρά , όπως και οι εναλλακτικές απαντήσεις για κάθε ερώτηση. Με τον τρόπο αυτό αυξάνεται η αξιοπιστία και η εγκυρότητα των αποτελεσμάτων. Ο εξεταζόμενος μπορεί να δει , εκτός από την ερώτηση και τις εναλλακτικές απαντήσεις , το χρόνο που του απομένει για την ολοκλήρωση του διαγωνίσματος , τον αύξοντα αριθμό της ερώτησης και το συνολικό αριθμό των ερωτήσεων του διαγωνίσματος. Επίσης του δίνεται η δυνατότητα να εξέλθει από το σύστημα (logout).



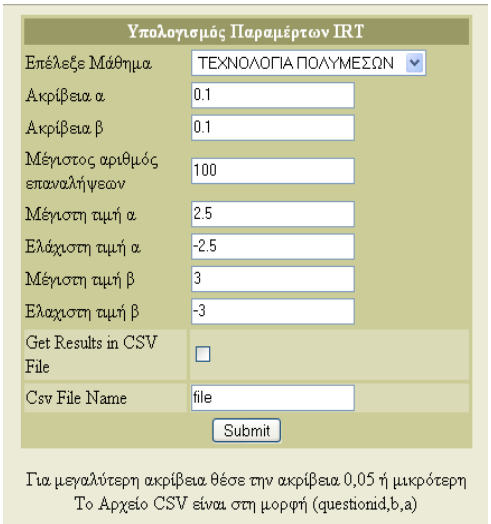
Group : 0		Group : 1	
AA	username	AA	username
1	c	1	a
2	b	2	a1
3	d	3	a2
4	e	4	a3
5	f	5	a4
6	g	6	a5
7	h	7	a6
8	i	8	a7
9	j	9	a8
10	b1	10	a9
11	b2		
12	b3		
13	b4		
14	b5		

Σχήμα 14 : *Η κύρια σελίδα της πλατφόρμας admin2 μέσω της οποίας γίνεται η ανάθεση των σπουδαστών σε ομάδες με κάποιες δοκιμαστικές εγγραφές σπουδαστών. Στο αριστερό πλαίσιο φαίνονται κάποιες γενικές πληροφορίες για την εξέταση και το μάθημα. Συγκεκριμένα φαίνεται ο αριθμός των σπουδαστών που έχουν εγγραφεί στην εξέταση , ο αριθμός των ομάδων και ο αριθμός των εξεταζόμενων σε κάθε ομάδα. Επίσης δίνεται η δυνατότητα στον υπεύθυνο της εξέτασης να επιλέξει κάποιο από τα μαθήματα τα οποία διαχειρίζεται και να εξέλθει από το σύστημα. Για κάθε μάθημα υπάρχουν 20 διαθέσιμες ομάδες οι οποίες μπορούν να ανατεθούν σε κάθε εξεταζόμενο. Στο κεντρικό πλαίσιο φαίνονται τα usernames των εξεταζόμενων στους οποίους έχει ανατεθεί κάποια ομάδα και σε παρένθεση η ομάδα που τους έχει ανατεθεί. Μέσω της κεντρικής φόρμας μπορεί να αλλάξει ο υπεύθυνος της εξέτασης την ομάδα κάποιου εξεταζόμενου. Στο δεξί πλαίσιο φαίνονται οι εξεταζόμενοι στους οποίους δεν έχει ανατεθεί κάποια ομάδα , και μέσω αυτής της φόρμας μπορεί να γίνει ανάθεση εξεταζόμενου σε κάποια από τις ομάδες. Στο κάτω μέρος παρουσιάζονται αναλυτικά όλες οι ομάδες και οι εξεταζόμενοι που ανήκουν στην καθεμία*

Αφού πραγματοποιηθούν οι δοκιμαστικές εξετάσεις υπολογίζονται οι παράμετροι για το κάθε αντικείμενο. Έχουν υλοποιηθεί δύο εφαρμογές PHP για την εκτίμηση των παραμέτρων. Η estimate1 υπολογίζει τις παραμέτρους για το μοντέλο 1PL , μιας παραμέτρου , και η estimate2 τις παραμέτρους για το μοντέλο 2PL. Και στις δύο εφαρμογές δίνεται η δυνατότητα στον αναλυτή να δει τα προβληματικά δεδομένα (ερωτήσεις με αρνητική διακριτική ικανότητα , ερωτήσεις που έχουν απαντηθεί σωστά από όλους ή από κανένα από τους εξεταζόμενους , ή εξεταζόμενους που έχουν απαντήσει σωστά σε όλες ή καμία από τις ερωτήσεις) , τα οποία σημειώνονται με κόκκινη γραμματοσειρά στη σελίδα των αποτελεσμάτων . Τα προβληματικά δεδομένα αγνοούνται στους υπολογισμούς.

Επίσης του δίνεται η δυνατότητα να εξάγει τα δεδομένα σε ένα CSV αρχείο για ανάλυση μέσω εξωτερικών προγραμμάτων. Για να γίνει σωστή εκτίμηση των παραμέτρων πρέπει να υπάρχουν αρκετά δείγματα , διαφορετικά το σφάλμα εκτίμησης είναι μεγάλο ή δεν μπορεί να συγκλίνει η μέθοδος της εκτίμησης σε κάποιο αποτέλεσμα.

Στο σχήμα 14 φαίνεται η αρχική φόρμα για τον καθορισμό των παραμέτρων της σελίδας υπολογισμού των παραμέτρων του μοντέλου 2PL estimate2.php. Οι προκαθορισμένες τιμές είναι κάποιες τυπικές τιμές για τον υπολογισμό των παραμέτρων , αλλά δίνεται η δυνατότητα στο χρήστη να τις αλλάξει, για την επίτευξη μεγαλύτερης ακρίβειας.



Σχήμα 14 : *Η αρχική φόρμα για τον καθορισμό των παραμέτρων της σελίδας υπολογισμού των παραμέτρων του μοντέλου 2PL. Ο χρήστης μπορεί να εισάγει την ακρίβεια υπολογισμού που επιθυμεί , τις μέγιστες και ελάχιστες τιμές των παραμέτρων καθώς και το μέγιστο αριθμό επαναλήψεων της μεθόδου. Επιπλέον του δίνεται η δυνατότητα να εξάγει τα δεδομένα σε ένα csv αρχείο , το οποίο μπορεί να εισάγει στη βάση δεδομένων του C.A.T.*

Αφού γίνει η ανάλυση των δεδομένων και εισαχθούν τα αποτελέσματα στη βάση , ο εξεταστής έχει στη διάθεσή του μια βαθμονομημένη τράπεζα αντικειμένων , την οποία μπορεί να χρησιμοποιήσει για την πραγματοποίηση προσαρμοστικών εξετάσεων CAT.

Αυτό που αλλάζει στην προσαρμοστική εξέταση είναι ότι οι ερωτήσεις στις οποίες θα απαντήσει ο κάθε εξεταζόμενος επιλέγονται από τον υπολογιστή. Οι περιορισμοί που θέτονται από τον εξεταστή αφορούν το χρόνο της εξέτασης , τον αριθμό των συνολικών ερωτήσεων του διαγωνίσματος και το μέγιστο αριθμό των ερωτήσεων από κάθε θεματική ενότητα του μαθήματος σε ένα διαγώνισμα.

Ο χρόνος που απομένει στον κάθε εξεταζόμενο για την ολοκλήρωση της εξέτασης ανανεώνεται στη βάση μέσω AJAX και PHP για λόγους ασφαλείας. Ο χρόνος εμφανίζεται στον εξεταζόμενο μέσω javascipt. Αν απενεργοποιήσει στον περιηγητή του την javascript δεν μπορεί να δει το χρόνο που του απομένει , αλλά ο χρόνος αυτός ανανεώνεται στη βάση και η εξέταση τερματίζεται κανονικά εφόσον ο χρόνος που προβλεπόταν περάσει.

Δ.1. Αλγόριθμοι που χρησιμοποιήθηκαν για το C.A.T.

Η ικανότητα του εξεταζόμενου υπολογίζεται μετά από κάθε απάντησή του χρησιμοποιώντας τη μέθοδο μέγιστης πιθανοφάνειας (MLE). Εφόσον η εκτίμηση της ικανότητας είναι μεγαλύτερη του 3 αυτή θεωρείται ίση με 3 , ενώ αν είναι μικρότερη του -3 θεωρείται -3.

Για να αποφευχθεί η υπερβολική χρησιμοποίηση κάποιων αντικειμένων χρησιμοποιήθηκε ένας συνδυασμός μεθόδων. Για την επιλογή των πρώτων πέντε αντικειμένων ακολουθείται ο ακόλουθος αλγόριθμος.

- Υπολογίζεται η πληροφορία Fisher κάθε αντικειμένου και βρίσκεται το αντικείμενο με τη μέγιστη πληροφορία I_{max}
- Επιλέγονται πέντε το πολύ αντικείμενα με τη μεγαλύτερη πληροφορία. Ένα αντικείμενο επιλέγεται εφόσον η πληροφορία του βρίσκεται στο διάστημα [0.9 I_{max} , I_{max}]
- Βρίσκεται από τη βάση ο μέσος χρόνος απάντησης για το καθένα από τα παραπάνω αντικείμενα.
- Στο καθένα από αυτά αντιστοιχίζεται ένας αριθμός ο οποίος εξαρτάται από το μέγεθος της πληροφορίας του και το μέσο χρόνο που απαιτείται για την απάντησή του. Ο αριθμός αυτός είναι μεγαλύτερος όσο μεγαλύτερη είναι η πληροφορία του αντικείμενου και μικραίνει όσο περισσότερος είναι ο χρόνος που απαιτείται για την απάντησή του.
- Οι παραπάνω αριθμοί κανονικοποιούνται ώστε το άθροισμά τους να είναι μονάδα. Τα μέτρα που προκύπτουν από τις κανονικοποιήσεις αυτές είναι οι πιθανότητες επιλογής του συγκεκριμένου αντικειμένου.
- Επιλέγεται ένας τυχαίος αριθμός στο διάστημα [0,1] και σύμφωνα με αυτόν και τις πιθανότητες επιλογής των αντικειμένων επιλέγεται τυχαία το επόμενο αντικείμενο.

Με τον τρόπο αυτό η επιλογή των πρώτων πέντε αντικειμένων γίνεται τυχαία από τα πέντε το πολύ αντικείμενα με τη μεγαλύτερη πληροφορία. Τα αντικείμενα με μεγαλύτερη πληροφορία έχουν μεγαλύτερη πιθανότητα επιλογής , όπως και τα αντικείμενα που απαιτούν λιγότερο χρόνο για να απαντηθούν.

Για τις υπόλοιπες ερωτήσεις η επόμενη ερώτηση επιλέγεται σύμφωνα με το κριτήριο της μέγιστης πληροφορίας Fisher.

Για τη διαχείριση του ισοζυγίου του περιεχομένου της εξέτασης χρησιμοποιείται η μέθοδος MCCAT. Με τον τρόπο αυτό μπορεί να καθορίζεται εκ των προτέρων το ποσοστό των ερωτήσεων από την κάθε θεματική ενότητα που θα υπάρχει σε κάθε διαγώνισμα.

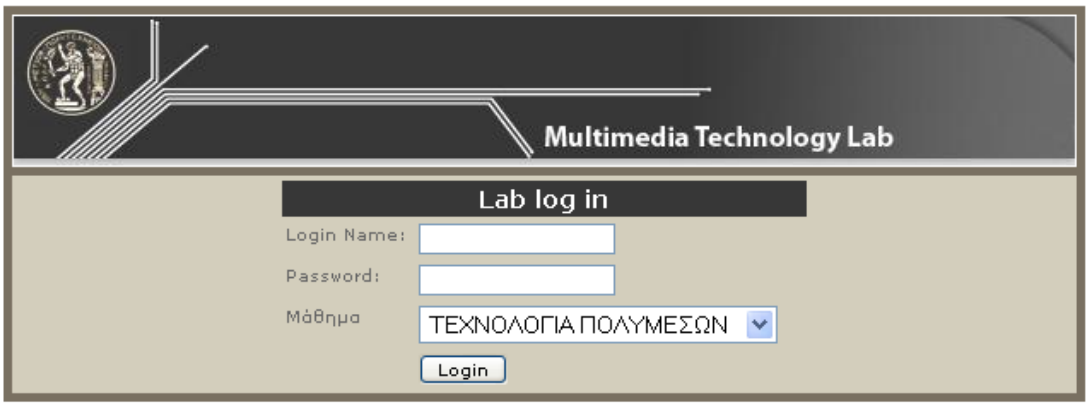
Δ.2. Use Cases

Δ.2.1. Use Cases για το script δοκιμαστικών διαγωνισμάτων.

Ο κάθε εξεταζόμενος συνδέεται στο σύστημα χρησιμοποιώντας τον προσωπικό του κωδικό.
Η αρχική φόρμα εισαγωγής φαίνεται στο σχήμα 15. Όταν συνδεθεί εμφανίζεται στον εξεταζόμενο μια ερώτηση από αυτές που έχουν ανατεθεί στην ομάδα του και δεν έχει απαντηθεί ήδη από τον εξεταζόμενο αυτό. Ο εξεταζόμενος επιλέγει μέσω μιας φόρμας μια από τις τέσσερις εναλλακτικές απαντήσεις που του δίνονται και υποβάλει την απάντησή του πατώντας το αντίστοιχο πλήκτρο.

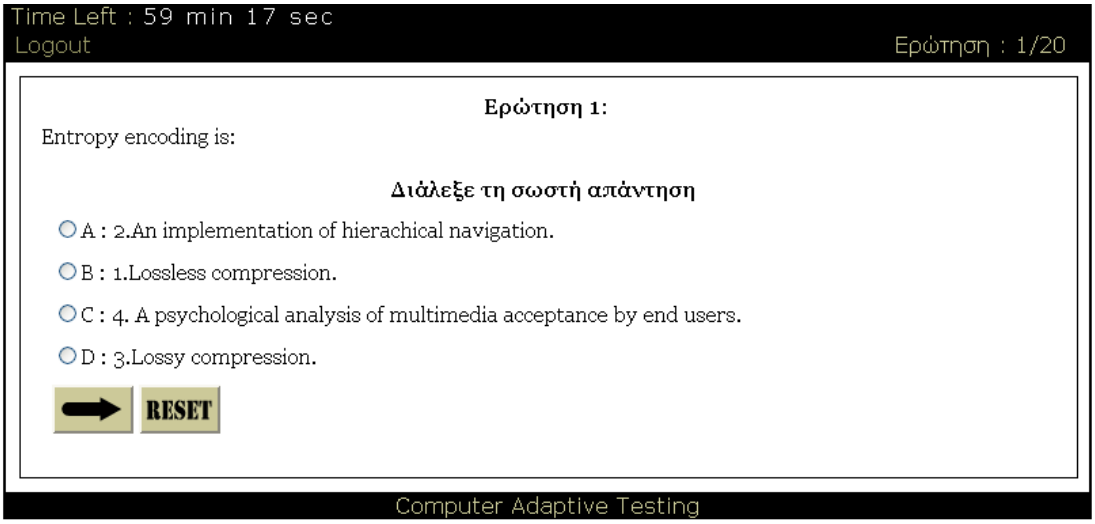
Στο σχήμα 16 φαίνεται η φόρμα που εμφανίζεται στον εξεταζόμενο για την απάντηση των ερωτήσεων.

Εφόσον ο εξεταζόμενος πατήσει το πλήκτρο ανανέωσης (refresh) ή επιστροφής (back) δεν αποθηκεύεται για τη συγκεκριμένη ερώτηση η απάντηση που έχει επιλέξει και εμφανίζεται η επόμενη ερώτηση , χωρίς να του δίνεται η επιλογή να απαντήσει στην ερώτηση αυτή στη συνέχεια της εξέτασης. Για το λόγο αυτό δε θα έπρεπε να χρησιμοποιούνται τα πλήκτρα αυτά ή οι συντομεύσεις τους.



Σχήμα 15 : *Η αρχική οθόνη για εισαγωγή στο περιβάλλον εξέτασης*

Ο κάθε εξεταζόμενος έχει τη δυνατότητα να εξέλθει από το σύστημα της εξέτασης περιορισμένες φορές. Αυτό πρέπει να γίνεται σε συνεννόηση με τους επιβλέποντες της εξέτασης για να διασφαλίζεται η σωστή χρήση του χρονομέτρου.



Σχήμα 16 : *Η κύρια οθόνη του περιβάλλοντος εξέτασης. Ο εξεταζόμενος μπορεί να δει το χρόνο που του απομένει για την ολοκλήρωση της εξέτασης , το συνολικό αριθμό των ερωτήσεων της εξέτασης, τον αύξοντα αριθμό της ερώτησης που του εμφανίζεται , την ερώτηση και τις εναλλακτικές απαντήσεις. Επιπλέον του δίνεται η δυνατότητα να εξέλθει από το περιβάλλον της εξέτασης(logout).*

Στην περίπτωση που ο εξεταζόμενος εξέλθει από το σύστημα (μέσω της επιλογής Logout ή κλείνοντας το πρόγραμμα περιήγησης που χρησιμοποιεί για να εξεταστεί οι πληροφορίες που έχει δώσει μέχρι τη στιγμή εκείνη έχουν αποθηκευτεί στη βάση και εφόσον συνεχίσει την εξέταση , αυτή συνεχίζεται από το σημείο στο οποίο είχε σταματήσει πριν εξέλθει από το σύστημα. Με τον τρόπο αυτό , αν προκύψει κάποιο πρόβλημα στη διάρκεια μιας εξέτασης μπορεί αυτή να συνεχιστεί από το σημείο που είχε σταματήσει , αφού διορθωθούν τα προβλήματα.

Σε περίπτωση που τελειώσει ο χρόνος της εξέτασης εμφανίζεται στον εξεταζόμενο ένα μήνυμα που τον πληροφορεί ότι η εξέταση έχει τελειώσει.

Δ.2.2 Use Cases για το script προσαρμοστικών διαγωνισμάτων.

Ο κάθε εξεταζόμενος συνδέεται στο σύστημα χρησιμοποιώντας τον προσωπικό του κωδικό.
Όταν συνδεθεί εμφανίζεται στον εξεταζόμενο μια ερώτηση οποία επιλέγεται από τον υπολογιστή σύμφωνα με το κριτήριο επιλογής αντικειμένου. Ο εξεταζόμενος επιλέγει μέσω μιας φόρμας μια από τις τέσσερις εναλλακτικές απαντήσεις που του δίνονται και υποβάλει την απάντησή του πατώντας το αντίστοιχο πλήκτρο. Οι σελίδες που εμφανίζονται στον εξεταζόμενο δεν έχουν καμία διαφορά ως προς την εμφάνιση με αυτές που εμφανίζονται στα δοκιμαστικά διαγωνίσματα , αυτό που αλλάζει είναι μόνο ο τρόπος επιλογής των ερωτήσεων.

Εφόσον ο εξεταζόμενος πατήσει το πλήκτρο ανανέωσης (refresh) ή επιστροφής (back) δεν αποθηκεύεται για τη συγκεκριμένη ερώτηση η απάντηση που έχει επιλέξει και εμφανίζεται στον εξεταζόμενο μήνυμα τερματισμού του διαγωνίσματος. Ο εξεταζόμενος δεν μπορεί να συνεχίσει το διαγώνισμα. Για το λόγο αυτό δε θα έπρεπε να χρησιμοποιούνται τα πλήκτρα αυτά ή οι συντομεύσεις τους.

Όπως και στα δοκιμαστικά διαγωνίσματα , υπάρχει περιορισμός στις φορές που μπορεί να συνδεθεί ένας εξεταζόμενος στο σύστημα για εξασφάλιση της σωστής χρήσης του χρονομέτρου. Επίσης εφόσον ο εξεταζόμενος εξέλθει από το σύστημα και εισέλθει δεύτερη φορά η εξέταση συνεχίζεται από το σημείο που είχε σταματήσει πριν αυτός να εξέλθει , όπως ακριβώς και στα δοκιμαστικά διαγωνίσματα.

Σε περίπτωση που τελειώσει ο χρόνος της εξέτασης εμφανίζεται στον εξεταζόμενο ένα μήνυμα που τον πληροφορεί ότι η εξέταση έχει τελειώσει.

Δ.3 Η βάση δεδομένων

Στη συγκεκριμένη υλοποίηση έχουν χρησιμοποιηθεί διχότομα αντικείμενα , η βάση δεδομένων όμως έχει σχεδιαστεί έτσι ώστε να μπορεί να χρησιμοποιηθεί στο μέλλον και για ανάλυση πολύτομων αντικειμένων, αν αυτό είναι επιθυμητό. Επίσης το σύστημα έχει σχεδιαστεί και υλοποιηθεί για το μοντέλο 2PL αλλά η βάση έχει δημιουργηθεί έτσι ώστε να μπορούν να χρησιμοποιηθούν και αντικείμενα του μοντέλου 3PL .
Το σχήμα της βάσης φαίνεται στο Σχήμα 15.

Ακολουθεί αναλυτική περιγραφή των πινάκων της βάσης και των πεδίων τους

Στον πίνακα **users** υπάρχουν οι πληροφορίες για τους εξεταζόμενους. Εκτός από τα πεδία για τα προσωπικά δεδομένα του καθενός υπάρχουν τα πεδία :
groupid όπου αποθηκεύεται ο κωδικός της ομάδας του εξεταζόμενου για τις δοκιμαστικές εξετάσεις
complete το πεδίο αυτό παίρνει την τιμή yes εφόσον ο εξεταζόμενος έχει τελειώσει το διαγώνισμα. Διαφορετικά έχει την τιμή No
last_login η ημερομηνία και η ώρα στην οποία έκανε Login για τελευταία φορά ο χρήστης αυτός στο σύστημα.
login_nos ο αριθμός των login που έχει κάνει ο χρήστης αυτός
timeleft ο χρόνος που απομένει στο χρήστη για την ολοκλήρωση της εξέτασης.
last_update_time. Το πεδίο αυτό χρησιμοποιείται για την ανανέωση του χρόνου από το σύστημα.

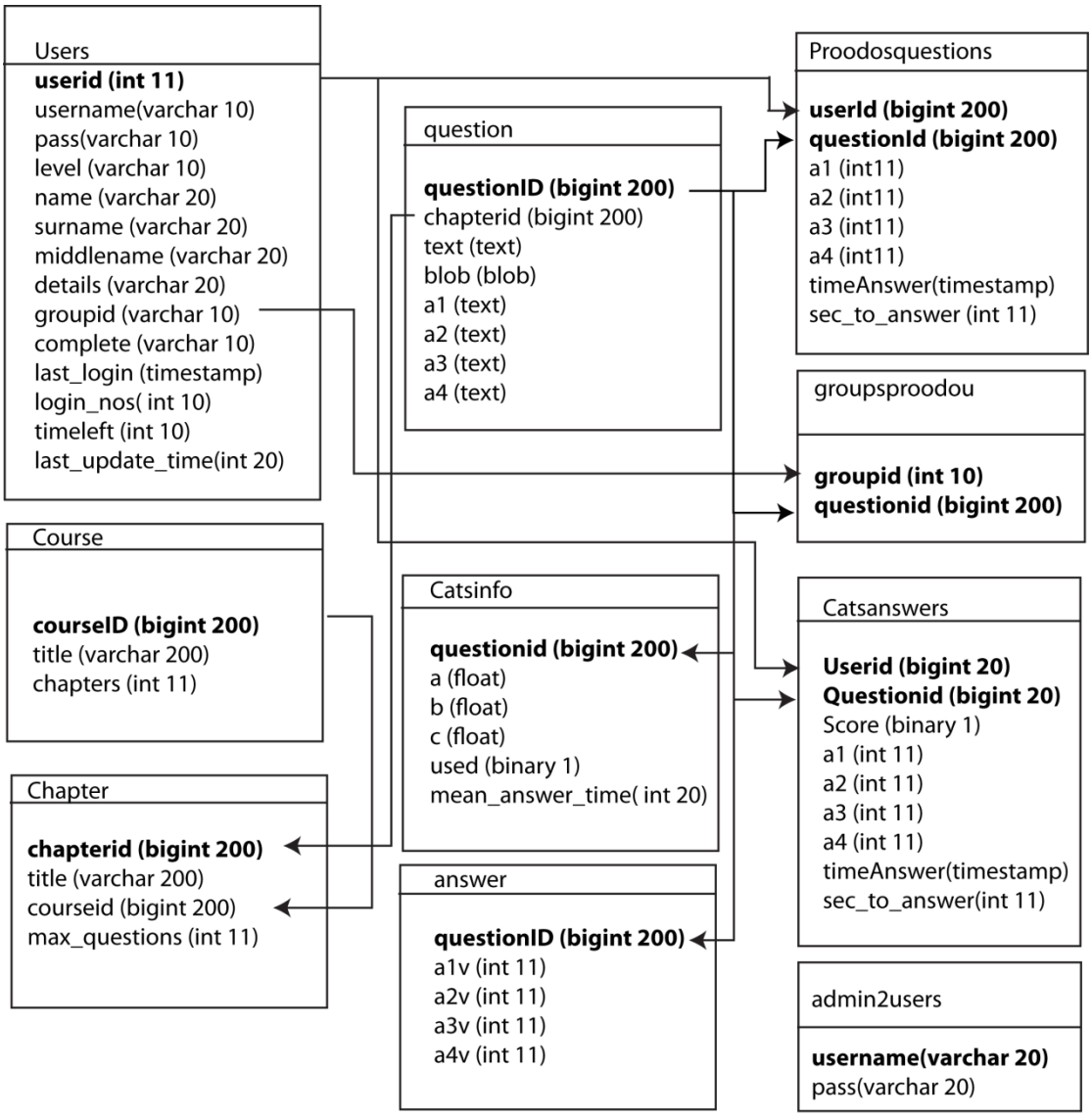
Όλα τα πεδία του users είναι κοινά για όλες τις εξετάσεις. Ο αριθμός των Login και ο χρόνος που υπολείπεται για την ολοκλήρωση του διαγωνίσματος πρέπει να ανανεώνεται πριν τη διεξαγωγή κάθε εξέτασης.

Πίνακας **question :** περιέχει πληροφορίες για τις ερωτήσεις
questionID : ο κωδικός της ερώτησης
chapterid : ο κωδικός του κεφαλαίου στο οποίο ανήκει η ερώτηση
text : το κείμενο της ερώτησης
blob : μια εικόνα που σχετίζεται με την ερώτηση
a1 ,a2 ,a3 ,a4 : το κείμενο των τεσσάρων εναλλακτικών απαντήσεων.

Πίνακας **answer**: περιέχει πληροφορίες για τις σωστές απαντήσεις
questionID : ο κωδικός της ερώτησης
a1v , a2v , a3v , a4v : η βαθμολογία της καθεμιάς από της εναλλακτικές απαντήσεις. Στην περίπτωση που υπάρχουν διχότομα αντικείμενα η σωστή απάντηση παίρνει την τιμή 1 ενώ οι άλλες την τιμή 0

πίνακας **Catsinfo** : περιέχει πληροφορίες για τη λειτουργία του CAT
questionid : ο κωδικός της ερώτησης
a , b , c : οι παράμετροι της ερώτησης
used : παίρνει την τιμή 1 αν ο εξεταστής επιθυμεί να χρησιμοποιηθεί η ερώτηση αυτή στο CAT. Διαφορετικά παίρνει την τιμή 0.
mean_answer_time : ο μέσος χρόνος απάντησης της ερώτησης αυτής.

Πίνακας **Course** : περιέχει πληροφορίες για τα μαθήματα
courseID : ο κωδικός του μαθήματος
title : ο τίτλος του μαθήματος
chapters : ο αριθμός των θεματικών ενοτήτων του μαθήματος



Σχήμα 15 : *Το σχήμα της βάσης*

Πίνακας **Chapter**: περιέχει πληροφορίες για τις θεματικές ενότητες των μαθημάτων
chapterid: ο κωδικός της θεματικής ενότητας
title : ο τίτλος της θεματικής ενότητας
courseid : ο κωδικός του μαθήματος στο οποίο ανήκει η θεματική ενότητα
max_questions : ο μέγιστος αριθμός ερωτήσεων από τη θεματική ενότητα αυτή που μπορούν να συμπεριληφθούν σε ένα διαγώνισμα CAT

Πίνακας **admin2users** περιέχει το username και το password για τους λογαριασμούς των διαχειριστών του συστήματος.

Πίνακας **Proodosquestions** περιέχει πληροφορίες σχετικά με τις απαντήσεις των εξεταζόμενων στα δοκιμαστικά διαγωνίσματα.
userId : ο κωδικός του χρήστη

questionId : ο κωδικός της ερώτησης
a1 , a2 , a3 , a4 : τα πεδία αυτά παίρνουν την τιμή 1 όταν ο χρήστης έχει επιλέξει την αντίστοιχει εναλλακτική απάντηση , διαφορετικά παίρνουν την τιμή 0.
timeAnswer : η ημερομηνία και η ώρα της απάντησης
sec_to_answer : ο χρόνος σε δευτερόλεπτα που χρειάστηκε ο χρήστης για να απαντήσει στη συγκεκριμένη ερώτηση.

Πίνακας **groupsproodou :** στον πίνακα αυτό αντιστοιχίζεται κάθε ομάδα εξεταζόμενων της δοκιμαστικής εξέτασης με τις ερωτήσεις που θα της ζητηθεί να απαντήσει.

Πίνακας **Catsanswers :** περιέχει πληροφορίες σχετικά με τις απαντήσεις των εξεταζόμενων στο CAT
Userid : ο κωδικός του εξεταζόμενου
Questionid : ο κωδικός της ερώτησης
Score : η βαθμολογία του εξεταζόμενου σε αυτό το αντικείμενο. Το πεδίο αυτό είναι μηδέν αν έχει δώσει λανθασμένη απάντηση ή 1 αν έχει δώσει σωστή.
a1 , a2 , a3 , a4 : τα πεδία αυτά παίρνουν την τιμή 1 αν ο εξεταζόμενος έχει επιλέξει την αντίστοιχη εναλλακτική απάντηση. Διαφορετικά παίρνουν την τιμή μηδέν.
timeAnswer : η ημερομηνία και η ώρα της απάντησης
sec_to_answer: ο χρόνος σε δευτερόλεπτα που χρειάστηκε ο χρήστης για να απαντήσει στη συγκεκριμένη ερώτηση.

E . References

- Baker Frank (1987)
  Item Parameter Estimation Under the One-, Two-, and Three-Parameter Logistic models
  Applied Psychological Measurement 1987; 11; 111

- Baker Frank (2001)
  The basics of item response theory
  ERIC Clearinghouse on Assessment and Evaluation

- Brown, J. S., & Burton, R. R. (1978). Diagnostic models
  for procedural bugs in basic mathematical skills.
  Cognitive Science, 2, 155-192.

- Brown, J. S., & VanLehn, K. (1980).
  Repair theory: A generative theory of bugs in procedural skills.
  Cognitive Science, 4, 379-426.

- Chen Shu-Ying , Ankenmann Robert D., Hua-Hua Chang 2000
  A comparison of item selection Rules at the early stages of computerized adaptive testing
  Applied Psychological Measurement 2000; 24; 241

- Cheng Ying & Chang Hua Hua 2007
  Two-Phase Item Selection Procedure for Flexible Content Balancing in CAT
  Applied Psychological Measurement 2007; 31; 467

- Georgiadou Elissavet G., Triantafillou Evangelos, and Economides Anastasios A. 2007
  A Review of Item Exposure Control Strategies for Computerized Adaptive Testing Developed from 1983 to 2005
  The Journal of Technology, Learning, and Assessment
  Volume 5, Number 8 · May 2007

- Lane, S., Stone, C. A., & Hsu, H. (1990, April).
  Diagnosing students' errors in solving algebra word problems.
  Paper presented at the annual meeting of the National Council on Measurement in Education, Boston.

- Leung, C.-K., Chang, H.-H., & Hau, K.-T.(2003)
  Computerized Adaptive Testing: A Comparison of Three Content Balancing Methods
  The Journal of Technology, Learning, and Assessment
  Volume 2, Number 5 · December 2003

- Levine, M., & Drasgow, F. (1983).
  The relation between incorrect option choice and estimated ability.

Educational and Psychological Measurement, 43, 675-685.

- Nedelsky, L. (1954).
  Ability to avoid gross error as a measure of achievement. Educational and Psychological Measurement, 14, 459-472.

- Tatsuoka, K. K. (1983).
  Rule space: An approach for dealing with misconceptions based on item response theory.
  Journal of Educational Measurement, 20, 345-354.

- Van der Linden , Wim J 1995
  Bayesian item selection in adaptive testing
  Annual Meeting of the Psychometric Society, Minneapolis MN, 1995

- Van der Linden , Wim J 1996.
  Bayesian Item Selection Criteria for adaptive testing
  Psychometrika, v63 n2 p201-16 Jun 1998

.....
Αναστάσιος Χ. Μαυρίδης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.