



Μεταπτυχιακή Διπλωματική Εργασία

**«Δημιουργία εργαλείου για την αξιολόγηση της
μαθηματικής ικανότητας, αποφοίτων δημοτικού, με χρήση
της Θεωρίας Απόκρισης Ερωτήματος»**

Ρίζου Αρσινόη Ειρήνη

Επιβλέπων καθηγητής: Μαυρίδης Δημήτριος

Ιωάννινα

Ιούλιος 2022



*Σχολή Επιστημών Αγωγής
Πανεπιστήμιο Ιωαννίνων*

Η Μεταπτυχιακή Διπλωματική Εργασία

«Δημιουργία εργαλείου για την αξιολόγηση της μαθηματικής ικανότητας, αποφοίτων δημοτικού, με χρήση της θεωρίας Απόκρισης Ερωτήματος»

υποβάλλεται από την

Αρσινόη Ειρήνη Ρίζου

στην Εξεταστική Επιτροπή του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης, ως μέρος των υποχρεώσεων για την απόκτηση του

**ΔΙΠΛΩΜΑΤΟΣ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΔΙΔΑΚΤΙΚΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΜΑΘΗΣΗΣ ΤΩΝ ΦΥΣΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ**

Εξεταστική επιτροπή:

- **Μαυρίδης Δημήτριος**, Αναπληρωτής Καθηγητής, Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Κώτσης Κωνσταντίνος**, Καθηγητής, Παιδαγωγικό Τμήμα Δημοτικής, Πανεπιστήμιο Ιωαννίνων
- **Τάτσης Κωνσταντίνος**, Αναπληρωτής Καθηγητής, Παιδαγωγικό Τμήμα Δημοτικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Στην οικογένειά μου.
Ευαγγελία, Σταύρος & Λάμπρος

ΕΥΧΑΡΙΣΤΙΕΣ

Με την παρούσα διπλωματική εργασία ολοκληρώνονται οι σπουδές μου στο μεταπτυχιακό πρόγραμμα σπουδών του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης, και επιθυμώ να δοθεί ιδιαίτερα μνεία σε όσους συνέβαλαν στον σκοπό αυτό.

Στις σπουδές μου, ήταν καθοριστική η συμβολή του επιβλέποντά μου κ. Μαυρίδη Δημήτριου, Αναπληρωτή Καθηγητή του τμήματος, ο οποίος με την επιστημονικά τεκμηριωμένη και συμβουλευτική καθοδήγηση που μου προσέφερε σε όλα τα στάδια εκπόνησης της μεταπτυχιακής εργασίας αλλά και με τις ενδεδειγμένες επικοινωνιακές παρατηρήσεις του, κατακτώ την περάτωση της εργασίας αυτής.

Επιπροσθέτως, θα ήθελα να ευχαριστήσω θερμά τον Αναπληρωτή Καθηγητή του τμήματος κ. Τάτση για την προθυμία, την ειλικρινή διάθεση και τις ουσιαστικές γνώσεις που αποκόμισα για την ορθή δόμηση μαθηματικών ερωτηματολογίων.

Εν κατακλείδι, δεν θα μπορούσα να μην αναφερθώ στους συναδέλφους των Γυμνασίων (7ο Γυμνάσιο Ιωαννίνων, 2ο Γυμνάσιο Ανατολής, Γυμνάσιο Λογγάδων) που με περίσσεια ενσυναίσθηση αποδέχτηκαν την πρόσκληση και συνέβαλαν με το να λάβουν οι μαθητές τους τα ερωτηματολόγια και να είμαστε σε θέση να συλλέξουμε και να αναλύσουμε τα δεδομένα.¹

¹ Η συλλογή και η επεξεργασία των δεδομένων προσωπικού χαρακτήρα που υποβάλλονται πραγματοποιείται σύμφωνα με τα οριζόμενα στις διατάξεις του Ν.4624/19 και του Κανονισμού (ΕΕ)2016/2019. Το Πανεπιστήμιο Ιωαννίνων συλλέγει και επεξεργάζεται τα δεδομένα προσωπικού χαρακτήρα αποκλειστικά στο πλαίσιο της υλοποίησης του σκοπού της παρούσας διαδικασίας. Για το χρονικό διάστημα που τα προσωπικά δεδομένα θα παραμείνουν στη διάθεση του Πανεπιστημίου Ιωαννίνων το υποκείμενο έχει τη δυνατότητα να ασκήσει τα δικαιώματά του σύμφωνα με τους όρους του Γενικού Κανονισμού Προστασίας Δεδομένων Προσωπικού Χαρακτήρα 2016/679 (Ε.Ε.) και τα οριζόμενα στα άρθρα 34 και 35 Ν. 4624/2019.

ΠΕΡΙΕΧΟΜΕΝΑ

Αφιέρωση	2
Ευχαριστίες	3
Κατάλογος Σχημάτων	iii
Κατάλογος Πινάκων	iv
Περίληψη	v
Extended Abstract	vi
ΚΕΦΑΛΑΙΟ 1 Εισαγωγή	2
1.1 Στόχοι.....	2
1.2 Δομή της Διατριβής	2
ΚΕΦΑΛΑΙΟ 2 Θεωρητικό υπόβαθρο	4
2.1 Κλασική Θεωρία Δοκιμών.....	4
2.1.1 Ανασκόπηση θεωρίας	4
2.1.2 Δομικά στοιχεία της θεωρίας.....	6
2.2 Δυνατά και Αδύναμα Σημεία της.....	7
2.2.1 Δυνατά Σημεία.....	7
2.2.2 Αδυναμία.....	7
2.3 Θεωρία Απόκρισης Ερωτήματος	8
2.3.1 Θεωρητικό Υπόβαθρο.....	9
2.3.2 Η αναμόρφωση των κανόνων και η διαφοροποίηση.....	11
2.4 Τα μοντέλα ανάλυσης.....	12
2.5 Κριτική Ανασκόπηση.....	18
ΚΕΦΑΛΑΙΟ 3 Ερωτηματολόγιο	23
3.1 Δόμηση ενός ερωτηματολογίου.....	23
3.2 Ερωτηματολόγιο μαθητών.....	27
3.3 Ερωτηματολόγιο Φοιτητών.....	34
ΚΕΦΑΛΑΙΟ 4 Αποτελέσματα - Συμπεράσματα	42
4.1 Αποτελέσματα.....	43

4.1.1	Μοντέλο μιας παραμέτρου	46
4.1.2	Μοντέλο δύο παραμέτρων	49
4.1.3	Μοντέλο τριών παραμέτρων.....	51
4.2	Σύγκριση μοντέλων.....	53
4.2.1	Επιλογή μοντέλου-μαθητές.....	53
4.2.2	Επιλογή Μοντέλου-φοιτητές	54
4.3	Γραφικές.....	62
4.3.1	Χαρακτηριστικές Καμπύλες Ερωτημάτων	62
4.3.2	Συνάρτηση Πληροφορίας Ερωτήματος	65
4.4	Συμπεράσματα	67
4.4.1	Μελλοντική έρευνα.....	67
	Βιβλιογραφία	69
	Σύντομο Βιογραφικό	80

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

<i>Εικόνα 1 Η θέση των καμπυλών διαφοροποιείται ανάλογα με τη δυσκολία τους και το επίπεδο ικανότητας των εξετασθέντων.....</i>	<i>10</i>
<i>Εικόνα 2 Χαρακτηριστικές Καμπύλες Στοιχείων που αντιστοιχούν στα διαφορετικά στοιχεία του μέτρου είναι παράλληλες κατά μήκος της κλίμακας ικανότητας.....</i>	<i>14</i>
<i>Εικόνα 3 Χαρακτηριστικές Καμπύλες που τέμνονται</i>	<i>16</i>
<i>Εικόνα 4 Σιγμοειδείς Καμπύλες</i>	<i>17</i>
<i>Εικόνα 5 Βρίσκοντας την ικανότητα με την Μέγιστη Πιθανοφάνειας</i>	<i>43</i>
<i>Εικόνα 6 Η "μεγάλη" εικόνα.....</i>	<i>45</i>
<i>Εικόνα 7 Χαρακτηριστικές Καμπύλες Ερωτημάτων (Μαθητές)</i>	<i>62</i>
<i>Εικόνα 8 Χαρακτηριστικές Καμπύλες Ερωτημάτων (Φοιτητές)</i>	<i>64</i>
<i>Εικόνα 9 Συνάρτηση Πληροφορίας (Μαθητές)</i>	<i>66</i>
<i>Εικόνα 10 Συνάρτηση Πληροφορίας (Φοιτητές).....</i>	<i>66</i>

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Σύγκριση Μεθόδων	20
Πίνακας 2 Δείγματα	43
Πίνακας 3 Ποσοστό σωστών απαντήσεων στις κοινές ερωτήσεις.....	44
Πίνακας 4 Συχνότητα Βαθμολογίας Μαθητές	46
Πίνακας 5 Συχνότητα Βαθμολογίας Φοιτητές	46
Πίνακας 6 Αποτελέσματα 1PL (μαθητές)	47
Πίνακας 7 Αποτελέσματα 1PL (φοιτητές)	48
Πίνακας 8 Αποτελέσματα 2PL (μαθητές)	49
Πίνακας 9 Αποτελέσματα 2PL (φοιτητές)	50
Πίνακας 10 Αποτελέσματα 3PL (μαθητές)	51
Πίνακας 11 Αποτελέσματα 3PL (φοιτητές)	52
Πίνακας 12 Επιπλέον πληροφορία από το επιλεχθέν μοντέλο.....	55
Πίνακας 13 Επιπλέον πληροφορία από το επιλεχθέν μοντέλο.....	57
Πίνακας 14 Αποτελέσματα X^2 test.....	60
Πίνακας 15 Αποτελέσματα X^2 test	60
Πίνακας 16 Σωστές απαντήσεις στις ερωτήσεις με αρνητικό <i>discrimination</i>	63
Πίνακας 17 Σωστές απαντήσεις στις ερωτήσεις με αρνητικό <i>discrimination</i>	64

ΠΕΡΙΛΗΨΗ

Ρίζου Αρσινόη Ειρήνη, Π.Μ.Σ. στη Διδακτική και τεχνολογίες μάθησης των Φυσικών Επιστημών, Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Σχολή Επιστημών Αγωγής, Πανεπιστήμιο Ιωαννίνων, Ιούνιος 2022

Θέμα: «Δημιουργία εργαλείου για την αξιολόγηση της επίδοσης αποφοίτων δημοτικού με χρήση της θεωρίας Απόκρισης Ερωτήματος»

Επιβλέπων: Μαυρίδης Δημήτριος, Αναπληρωτής Καθηγητής

Η αξιολόγηση της πραγματικής βασικής μαθηματικής ικανότητας των μαθητών αποτελεί ένα φλέγον ζήτημα καθώς ένα απλό διαγώνισμα με την κλασική ισοδύναμη βαθμολόγηση-αναντίρρηση δεν μπορεί να αντιστοιχήσει στην πραγματική ικανότητα τους παρότι είναι ο πιο σύνηθες τρόπος για την αξιολόγηση της ικανότητας στα Μαθηματικά. Η παρούσα διπλωματική, έχει στόχο στην δόμηση ενός σταθμισμένου εργαλείου μέτρησης της πραγματικής απόδοσης στα Μαθηματικά με την εφαρμογή των μοντέλων (μιας, δύο και τριών παραμέτρων) της Θεωρίας Απόκρισης Ερωτήματος και την ανάπτυξη κώδικα μέσω του ευρέως διαδεδομένου στατιστικού περιβάλλοντος R και το πακέτο LTM (Latent Trait Models). Επιπλέον, επιχειρείται ταυτόχρονα και μια σύγκριση των απαντήσεων των μαθητών με αυτές μελλοντικών δασκάλων που θα κληθούν να διδάξουν μαθηματικά στο Δημοτικό.

Λέξεις Κλειδιά: Θεωρία Απόκρισης Ερωτήματος, Μαθηματική Ικανότητα, Ερωτηματολόγιο, Στατιστικό περιβάλλον R

EXTENDED ABSTRACT

Rizou Arsinoi Eirini, M.Sc. in Didactics and Learning Techniques in Natural Sciences, School of Pedagogical, Pedagogical Departments of Primary Education, University of Ioannina, Greece, June 2022

Thesis Title: Creating a tool for evaluating the performance of graduate primary school students using Item Response Theory

Advisor: Mavridis Dimitrios Associate Professor

Assessing students' true basic math ability is a burning issue as a simple quiz with the classic equivalent grading-undeniably cannot match their actual ability although it is the most common way to assess math ability. The present dissertation aims to build a weighted tool for measuring actual performance in Mathematics by applying the 1PL, Rasch, 2PL and 3PL models of the Question Answer Theory and developing code through the widely used statistic R environment and the LTM (Latent Trait Models) package. In addition, a comparison is made at the same time with the students' answers with those of future teachers who will be called to teach mathematics in primary school.

Keywords: Item Response Theory, Mathematical ability, questionnaires, R software environment for statistical computing

1.1 Στόχοι

1.2 Δομή της Διατριβής

1.1 Στόχοι

Το πρώτο κεφάλαιο αποτελεί μία εισαγωγή στο αντικείμενο της διατριβής, η οποία είναι απαραίτητη για να παρουσιαστούν οι κύριοι στόχοι της, οι οποίοι είναι:

- Να δομηθεί ερωτηματολόγιο που αποτελεί έρεισμα για τη συλλογή των δεδομένων και εφαλτήριο για την επικείμενη αξιολόγηση ώστε να λειτουργήσει τελικώς ως σταθμισμένο εργαλείο μέτρησης της απόδοσης.
- Να αποτιμηθούν τα αποτελέσματα της έρευνας και να καταστεί εφικτή μελλοντική έρευνα επ' αυτής.

Συνεπώς, απαιτείται σε πρώτο στάδιο να αναλυθεί η σχετική με την αξιολόγηση-θεωρίας επίδοσης και ικανότητας των μαθητών στο μάθημα των Μαθηματικών να παραχθεί σχετικός κώδικας για την ανάλυση των δεδομένων αυτών.

1.2 Δομή της Διατριβής

Η διατριβή περιέχει 4 κεφάλαια για τα οποία δίνεται μία συνοπτική μα ενδελεχής περιγραφή.

- Το κεφάλαιο που ακολουθεί επι του τρέχοντος, είναι το δεύτερο κεφάλαιο που απαρτίζεται από την εκτενή αναφορά στο θεωρητικό υπόβαθρο της παρούσας διπλωματικής εργασίας. Παρουσιάζεται η Κλασική Θεωρία Δοκιμών που ήταν και η πιο επικρατέστερη, οι λόγοι που την έκαναν δημοφιλή αλλά και οι αδυναμίες της που ανέδειξαν την ανάγκη για την ανάπτυξη της Θεωρίας Απόκρισης

Ερωτήματος που είναι και το αντικείμενο του παρόντος εγγράφου. Αναλύονται τα πλεονεκτήματα της μεθόδου, τα μοντέλα που εφαρμόζονται, καθώς και μια αναλυτική σύγκριση των δύο προσεγγίσεων.

- Ακολουθεί το τρίτο κεφάλαιο που εμπεριέχει δομικά στοιχεία για την κατασκευή ενός υποδειγματικού ερωτηματολογίου καθώς και τα ερωτηματολόγια που διανεμήθηκαν στους φοιτητές και τους μαθητές στο πλαίσιο της μεταπτυχιακής διατριβής.
- Το τέταρτο και τελευταίο κεφάλαιο αφορά στην ανάλυση, την ερμηνεία και τα αποτελέσματα που ελήφθησαν από την παραγωγή κώδικα. Επιπλέον, εμπεριέχεται πρόταση-προθήκη για μελλοντική εργασία με αφορμή τα αποτελέσματα.
- Μετά την βιβλιογραφία ακολουθούν τα Παραρτήματα που περιέχουν α) το πακέτο-βιβλιοθήκη που εφαρμόστηκε και σχετικές πληροφορίες επ' αυτού β) συγκριτική ανασκόπηση των κοινών ερωτήσεων μαθητών-φοιτητών

ΚΕΦΑΛΑΙΟ 2

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

-
- 2.1 Κλασική Θεωρία Μέτρησης
 - 2.2 Δυνατά και Αδύναμα Σημεία της Μεθόδου
 - 2.4 Θεωρία Απόκρισης Ερωτήματος
 - 2.5 Ανάλυση
 - 2.6 Κριτική Ανασκόπηση
-

Υπάρχουν 2 κύριοι τρόποι ανάλυσης των αποτελεσμάτων ενός τεστ/εργαλείου/ερωτηματολογίου. Ο κλασικός τρόπος δηλαδή η Κλασική Θεωρία των Τεστ (Classical Test Theory) και η Θεωρία Απόκρισης Ερωτήματος (Item Response Theory)

2.1 Κλασική Θεωρία Δοκιμών

2.1.1 Ανασκόπηση θεωρίας

Η Κλασική Θεωρία Μέτρησης (CTT) είναι το σύνολο των εννοιών και των μεθόδων που έχουν αποτελέσει τη βάση για την ψυχομετρική αξιολόγηση. Η κλασική θεωρία δοκιμών υποστηρίζει ότι οι παρατηρούμενες βαθμολογίες είναι η αθροιστική συνάρτηση των αληθινών βαθμολογιών και των όρων σφάλματος. Οι πραγματικές βαθμολογίες είναι η ιδανική αξία μιας κατασκευής σε ένα συγκεκριμένο στοιχείο ή κατάσταση. Ο όρος σφάλματος, είναι η επίδραση παραγόντων που είναι ξένοι στο κατασκευάσμα ενδιαφέροντος. Οι όροι σφάλματος θεωρούνται ανεξάρτητοι (ή ασύνδετοι με) τις πραγματικές βαθμολογίες. Η ανάλυση της αξιοπιστίας μιας βαθμολογίας μπορεί να επιτευχθεί με χειρισμό παραγόντων που πιστεύεται ότι επηρεάζουν τον όρο σφάλματος.

Πιο απλά, η CTT, που μερικές φορές αποκαλείται μοντέλο αληθινής βαθμολογίας, είναι τα μαθηματικά πίσω από τη δημιουργία και την απάντηση σε τεστ της κλίμακας μέτρησης. Ο στόχος της είναι να βελτιώσει τις δοκιμές, ιδιαίτερα την αξιοπιστία και την εγκυρότητα των δοκιμών. Η αξιοπιστία (McDonald, 1999) αφορά το πόσο καλά μετριέται ένα κατασκευάσμα, σε αντίθεση με την εγκυρότητα, η οποία αφορά κυρίως τη χρησιμότητα του τεστ ή του μέτρου για διάφορους πιθανούς σκοπούς. Η αξιοπιστία συνεπάγεται συνέπεια: εάν για παράδειγμα κάποιος επαναλαμβάνει την δοκιμασία πέντε φορές, θα πρέπει να έχει περίπου τα ίδια αποτελέσματα κάθε φορά, ενώ ένα τεστ είναι έγκυρο εάν μετρά αυτό που "υποτίθεται".

Η Κλασική Θεωρία Δοκιμών υποθέτει ότι κάθε άτομο έχει μια έμφυτη αληθινή βαθμολογία και μπορεί να συνοψιστεί με μια εξίσωση που οι μεταβλητές καθορίζονται ως ακολούθως

$$X = T + E,$$

όπου

- X είναι μια παρατηρούμενη βαθμολογία,
- T είναι η πραγματική βαθμολογία
- E είναι τυχαίο σφάλμα.
- Η αξιοπιστία είναι

$$r = \frac{\text{διακύμανση } T}{\text{διακύμανση } X}$$

Επειδή το σφάλμα και η αξιοπιστία αντιστοιχούν άμεσα το ένα στο άλλο, ο τύπος αξιοπιστίας που αξιολογούμε για ένα μέτρο εξαρτάται από τον τύπο του σφάλματος που επιδιώκουμε να αξιολογήσουμε. Όταν το σφάλμα μέτρησης σε ένα μέτρο είναι ανησυχητικό, επιδιώκουμε να εξακριβώσουμε πόση μεταβλητότητα στις βαθμολογίες μπορεί να αποδοθεί στην πραγματική μεταβλητότητα σε αντίθεση με το σφάλμα. Το σφάλμα μέτρησης σε ένα μέτρο εκδηλώνεται ως αποτέλεσμα της δειγματοληψίας περιεχομένου και της ετερογένειας της συμπεριφοράς του δείγματος. Η δειγματοληψία περιεχομένου αναφέρεται στη δειγματοληψία στοιχείων που συνθέτουν το μέτρο. Εάν τα στοιχεία του δείγματος προέρχονται από τον ίδιο τομέα, το σφάλμα μέτρησης σε ένα μέτρο θα είναι μικρότερο. Η ετερογένεια συμπεριφοράς μπορεί να οδηγήσει σε αύξηση του σφάλματος μέτρησης όταν τα στοιχεία αντιπροσωπεύουν διαφορετικούς τομείς συμπεριφορών. Άλλες πηγές σφαλμάτων μέτρησης σε μια δοκιμή μπορεί να προκύψουν, όπως εικασίες, λάθη και σφάλματα βαθμολογίας.

Για να καταστεί πιο σαφές, έστω το εξής παράδειγμα, ας υποθέσουμε ότι γνωρίζετε ακριβώς το 70% του συνόλου της ύλης σε ένα εξεταζόμενο μάθημα. Αυτή είναι και η πραγματική βαθμολογία σας (T). Ένα τέλειο διαγώνισμα (που δεν υφίσταται) στο τέλος του εξαμήνου θα πρέπει ιδανικά να αντικατοπτρίζει αυτό το πραγματικό σκορ. Στην πραγματικότητα, είναι πι-

θανό να σημειώσετε περίπου 65% έως 75%. Η απόκλιση 5% από την πραγματική βαθμολογία σας είναι το σφάλμα (E). Τα σφάλματα, θεωρείται, ότι κατανέμονται κανονικά με μέσο όρο μηδέν. Υποθετικά, εάν κάνατε το τεστ άπειρες φορές, η παρατηρούμενη βαθμολογία σας θα πρέπει να είναι ίση με την πραγματική σας βαθμολογία.

2.1.2 Δομικά στοιχεία της θεωρίας

Η κλασική θεωρία δοκιμών είναι μια συλλογή πολλών στατιστικών στοιχείων, συμπεριλαμβανομένης της μέσης βαθμολογίας, της δυσκολίας του ερωτήματος και της αξιοπιστίας του τεστ (Cohen and Manion, 1998)

- **Συσχέτιση:** δείχνει πώς δύο μεταβλητές X και Y σχετίζονται μεταξύ τους. Χρησιμοποιούνται διαφορετικά μέτρα για διαφορετικούς τύπους δοκιμών. Για παράδειγμα, ένα τεστ με διχοτομική βαθμολογία (π.χ. απαντήσεις ναι/όχι) θα συσχετιστεί με τη συσχέτιση σημείων, ενώ ένα τεστ με πολυτομική βαθμολογία (ένα με πολλαπλές απαντήσεις) θα βαθμολογηθεί με τον Συντελεστή Συσχέτισης Pearson.
- **Συνδιακύμανση:** είναι ένα μέτρο του πόσο ποικίλλουν δύο τυχαίες μεταβλητές μαζί. Είναι παρόμοιο με τη διακύμανση, αλλά όπου η διακύμανση λέει πώς ποικίλλει μια μεμονωμένη μεταβλητή, η συνδιακύμανση λέει πώς ποικίλλουν δύο μεταβλητές μαζί.
- **Δείκτης Διακρίσεων:** η ικανότητα του τεστ να κάνει διάκριση μεταξύ διαφορετικών επιπέδων μάθησης ή άλλης έννοιας ενδιαφέροντος. Ένας υψηλός δείκτης διάκρισης υποδηλώνει ότι το τεστ είναι σε θέση να διαφοροποιήσει μεταξύ των επιπέδων.
- **Δυσκολία αντικειμένου:** μέτρο δυσκολίας ατομικής ερώτησης εξέτασης. Είναι το ποσοστό των εξεταζόμενων που απάντησαν σωστά από τον συνολικό αριθμό των εξεταζόμενων. Για παράδειγμα, μια βαθμολογία δυσκολίας αντικειμένου 89/100 σημαίνει ότι από τα 100 άτομα, τα 89 απάντησαν σωστά.
- **Συντελεστής αξιοπιστίας:** ένα μέτρο για το πόσο καλά μετρά το τεστ το επίτευγμα. Υπάρχουν διάφορες μέθοδοι για τον υπολογισμό του συντελεστή, όπως δοκιμή-επανελέγχος, παράλληλη ή εναλλακτική μορφή και εσωτερική ανάλυση, εμπειρικοί κανόνες για τα προτιμώμενα επίπεδα του συντελεστή.

- Διακύμανση δείγματος / Τυπική απόκλιση: Η διακύμανση του δείγματος και η τυπική απόκλιση δείγματος είναι μέτρα για την κατανομή των βαθμολογιών.
- Τυπικό Σφάλμα Μέτρησης: ένα μέτρο του πόσες βαθμολογίες μετρήσεων κατανέμονται γύρω από μια «αληθινή» βαθμολογία.

2.2 Δυνατά και Αδύναμα Σημεία της

2.2.1 Δυνατά Σημεία

Εν γένει, τα πιο σημαντικά πλεονεκτήματα-τουλάχιστον-πριν αρκετά χρόνια θεωρούνταν τα ακόλουθα

- η εξοικείωση του ερευνητικού και επιστημονικού κοινού με την μέθοδο
- ορισμένες στατιστικές δοκιμές CTT είναι διαθέσιμες συνήθως σε πολλά δημοφιλή στατιστικά πακέτα.
- Οι αδύναμες υποθέσεις που προηγούνται για τα δεδομένα από το CTT αναφέρονται συχνά ως πλεονέκτημα που κάνει το CTT πιο ευρέως εφαρμόσιμο, όπως είναι η εννοιολογική απλότητα του μοντέλου (δηλαδή, $X = T + E$)

2.2.2 Αδυναμία

Η CTT (Polit, 2006) έχει αρκετές αδυναμίες που οδήγησαν στην ανάπτυξη άλλων μοντέλων για τις βαθμολογίες των τεστ. Πρώτον, η έννοια της αξιοπιστίας εξαρτάται από την ομάδα που χρησιμοποιήθηκε για την ανάπτυξη του τεστ. Εάν η ομάδα έχει ένα ευρύ φάσμα δεξιοτήτων ή ικανοτήτων, τότε η αξιοπιστία θα είναι υψηλότερη από ό,τι εάν η ομάδα έχει ένα στενό εύρος δεξιοτήτων ή ικανοτήτων. Επομένως, η αξιοπιστία δεν είναι αμετάβλητη σε σχέση με το δείγμα των εξεταζόμενων και επομένως δεν αποτελεί χαρακτηριστικό της ίδιας της δοκιμής. Επιπλέον, ούτε τα κοινά μέτρα διάκρισης αντικειμένων (όπως η συσχέτιση στοιχείου-συνολικού) ή η δυσκολία αντικειμένου (ποσοστό απάντησης του αντικειμένου-σωστά). Επιπλέον θεωρεί ότι όλα τα ερωτήματα σε ένα τεστ συνεισφέρουν ισοδύναμα στη συνολική επίδοση του εξεταζόμενου και γίνεται η συνήθης υπόθεση ότι το τυπικό σφάλμα είναι το ίδιο για όσους κάνουν τεστ σε όλα τα επίπεδα ικανότητας είναι εσφαλμένη. Σε ορισμένες επεκτάσεις της CTT, αυτή η υπόθεση απορρίπτεται, (αλλά αυτές οι επεκτάσεις δεν είναι ευρέως γνωστές ή ευρέως χρησιμοποιούμενες), οι υποψήφιοι βαθμολογούνται είτε στο χαμηλό είτε στο υψηλό όριο του εύρους βαθμολογίας του τεστ. Στην ουσία του πράγματος, εφαρμόζοντας την CTT, σε ένα εύκολο τεστ ένας φοιτητής μετρίων δυνατοτήτων πλησιάζει

σε επίδοση έναν πολύ καλό φοιτητή ενώ σε ένα δύσκολο τεστ ένας μέτριος φοιτητής πλησιάζει σε επίδοση έναν πολύ αδύναμο φοιτητή.

Επιπλέον παρουσιάζει δυστοκία στο χειρισμό ορισμένων τυπικών προβλημάτων ανάπτυξης δοκιμών, οριζόντιας και κάθετης εξίσωσης. Το πρόβλημα της οριζόντιας εξίσωσης προκύπτει όταν κάποιος επιθυμεί να αναπτύξει ένα άλλο τεστ με τις ίδιες ιδιότητες (ή τουλάχιστον με μια γνωστή σχέση με) ένα υπάρχον τεστ. Για παράδειγμα, οι μαθητές που κάνουν εξετάσεις εισαγωγής σε ένα κολέγιο, θα πρέπει να λάβουν την ίδια βαθμολογία ανεξάρτητα από την έκδοση του τεστ που θα κάνουν. Η κάθετη εξίσωση περιλαμβάνει την ανάπτυξη μιας σειράς τεστ που μετρούν ένα ευρύ φάσμα ικανοτήτων ή δεξιοτήτων. Για παράδειγμα, ενώ θα μπορούσαμε να αναπτύξουμε εντελώς ανεξάρτητα τεστ αριθμητικής για κάθε τάξη του δημοτικού σχολείου, θα μπορούσαμε να τα συνδέσουμε με μια συνεχή κλίμακα μαθηματικών δεξιοτήτων για τις τάξεις Α' δημοτικού έως Στ' Δημοτικού. Ενώ η κάθετη και η οριζόντια εξίσωση δεν είναι αδύνατη στο CTT, είναι πολύ πιο απλή με τη θεωρία απόκρισης αντικειμένων (IRT) όπως θα δούμε στην επόμενη ενότητα.

Αυτά τα προβλήματα οφείλονται εν μέρει σε κάποια ασάφεια στη θεωρία (ο πληθυσμός προς δειγματοληψία συνήθως δεν λαμβάνεται υπόψη με καμία λεπτομέρεια στη θεωρία). Αλλά οφείλονται επίσης στην αποτυχία των περισσότερων συλλογών δεδομένων να πραγματοποιηθούν σε ένα τυχαίο δείγμα από οποιονδήποτε πληθυσμό, πόσο μάλλον σε έναν πληθυσμό που θεωρείται κατάλληλος για τη δοκιμή που ερευνάται. Στην πράξη, χρησιμοποιούνται δείγματα ευκολίας. Αυτά γενικά ταιριάζουν σε ορισμένα κριτήρια που καθορίζει ο ερευνητής, αλλά λαμβάνονται διαφορετικά καθώς είναι διαθέσιμα στον ερευνητή. Τέλος, η CTT, όπως είχε αρχικά σχεδιαστεί, δεν προοριζόταν ποτέ να αντιμετωπίσει ορισμένα από τα πρακτικά προβλήματα δοκιμών που περιγράφηκαν παραπάνω.

2.3 Θεωρία Απόκρισης Ερωτήματος

Ένας άλλος κλάδος της ψυχομετρικής θεωρίας είναι η Θεωρία Απόκρισης Ερωτήματος (Item Response Theory) και πρόκειται για μια από τις πιο ουσιαστικές εξελίξεις στον τομέα της εκπαιδευτικής και ψυχολογικής μέτρησης. Μερικές φορές αναφέρεται ως Θεωρία των Λανθανόντων Χαρακτηριστικών καθώς και ως θεωρία της ισχυρής πραγματικής επίδοσης ή ισχυρή αληθινή θεωρία βαθμολογίας ή η σύγχρονη (παρότι εφαρμόζεται από τις απαρχές του 1970) θεωρία νοητικών τεστ, επειδή το IRT είναι ένα πιο πρόσφατο σώμα θεωρίας και κάνει ισχυρότερες υποθέσεις σε σύγκριση με την Κλασική Θεωρία Δοκιμών. Αυτή η προσέγγιση που βασίζεται στην ανάλυση αντικειμένων (ερωτημάτων) εξετάζει την πιθανότητα να απαντηθούν σωστά ή λάθος συγκεκριμένα αντικείμενα. Σε αυτή την προσέγγιση, κάθε στοιχείο

σε ένα τεστ έχει τη δική του χαρακτηριστική καμπύλη στοιχείων που εξετάζει την πιθανότητα σωστής απάντησης ως συνάρτηση της δυσκολίας κάθε ερώτησης, της ικανότητας του ερωτώμενου και της ικανότητας της ερώτησης να ξεχωρίζει άτομα διαφορετικής ικανότητας (Li et al., 2012).

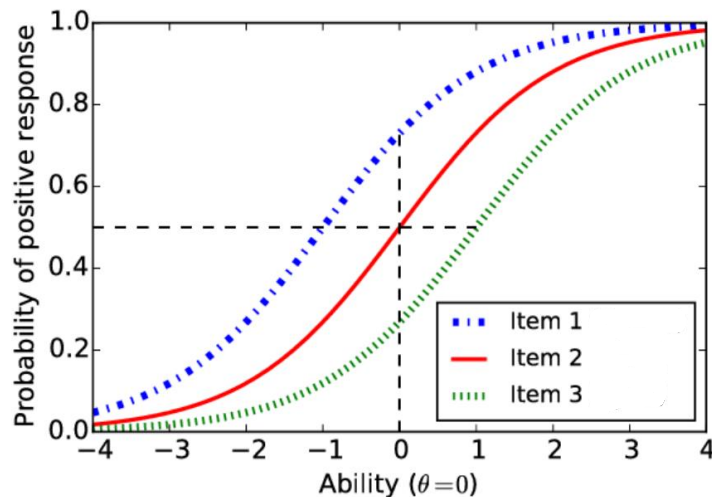
2.3.1 Θεωρητικό Υπόβαθρο

Έχει γνώμονα την ανάλυση των απαντήσεων των εξεταζόμενων στα επιμέρους ερωτήματα (items) που συνθέτουν ένα τεστ παρά στα συνολικά αποτελέσματα του. Η μέτρηση ενός χαρακτηριστικού εξαρτάται ταυτόχρονα από τις απαντήσεις των εξεταζόμενων και από τις ιδιότητες των σχετικών ερωτημάτων. Με στόχο τη βελτίωση των διαδικασιών μέτρησης των διαφόρων χαρακτηριστικών των ατόμων όπως γνώσεις και δεξιότητες, είναι πιο σύνθετη σε σύγκριση με την Κλασική. Ο κεντρικός άξονας της θεωρίας είναι ο προσδιορισμός του βαθμού της καταλληλότητας των ερωτημάτων που χρησιμοποιούνται σε μια δοκιμασία για να επιτευχθεί όσο γίνεται καλύτερα το επιδιωκόμενο αποτέλεσμα (Lord, 1980). Η αξιολόγηση της καταλληλότητας τους είναι μείζουσας σημασίας ζήτημα όχι μόνο για την κατασκευή και την αποτίμηση της αξίας ενός τεστ αλλά και για την οργάνωση των ερωτημάτων στις σχετικές τράπεζες θεμάτων καθώς και για τις διάφορες συγκριτικές διαχρονικές μελέτες αξιολόγησης, ή για τις αντίστοιχες μετα-αναλύσεις.

Η συγκεκριμένη θεωρία στηρίζεται στις εξής κυρίως παραδοχές

- α)** οι απαντήσεις των εξεταζόμενων στα ερωτήματα ενός τεστ αποτελούν παρατηρήσιμες μορφές συμπεριφοράς που αντανakλούν λανθάνουσες ικανότητές τους οι οποίες καθορίζουν την πιθανότητα που έχουν να απαντήσουν σωστά στα ερωτήματα τα οποία τους τίθενται παραδείγματος χάρη επίπεδο σχετικών γνώσεων και δεξιοτήτων στάσεις έναντι ορισμένων ζητημάτων νοητική ικανότητα και τα λοιπά. Η δυνατότητα αυτή (χαρακτηριστικό) συμβολίζεται συνήθως με το γράμμα θ και μπορεί να μετρηθεί με βάση μια κλίμακα λογιστικών μονάδων (logits) που έχουν μέσο όρο μηδέν και τυπική απόκλιση ίση με ± 1 . Οι συνήθεις τιμές που λαμβάνει το λανθάνον χαρακτηριστικό κυμαίνονται από -4 έως +4 αλλά είναι ίσως υπάρχουν και μεγαλύτερες τιμές αφού θεωρητικά η κλίμακα αυτή εκφράζει διάστημα που εκτείνεται από το μείον άπειρο έως το συν άπειρο.
- β)** για άτομα συγκεκριμένης ικανότητας, τα ερωτήματα είναι ανεξάρτητα, δηλαδή προϋποθέτει την ανεξαρτησία όταν λάβουμε υπόψιν την ικανότητα του μαθητή.
- γ)** οι παράμετροι δυσκολία διακριτικότητα παραμένουν οι ίδιες σε διαφορετικά δείγματα εξεταζόμενων. Συνεπώς, δεν απαιτείται τυχαία δειγματοληψία χρειάζονται όμως δείγματα που να διασφαλίζουν ικανό εύρος της μετρούμενης λανθάνουσας ικανότητας.

- δ) οι τιμές των παραμέτρων των ερωτημάτων και ο βαθμός της λανθάνουσας ικανότητας των εξεταζόμενων εκφράζονται στην ίδια μετρική κλίμακα και τοποθετούνται στον ίδιο οριζόντιο άξονα.
- ε) Η απάντηση του εξεταζόμενου σε ένα ερώτημα μπορεί να εκφραστεί ως λογιστική συνάρτηση (Item Response Function-I.R.F) που δηλώνει την πιθανότητα (P_i) ενός ατόμου με ορισμένο χαρακτηριστικό θ να απαντήσει σωστά σε ένα ερώτημα το οποίο έχει συγκεκριμένες ιδιότητες (παραμέτρους). Η συνάρτηση αυτή εκφράζεται με τη μορφή μιας σιγμοειδούς καμπύλης (Εικόνα 1) ονομάζεται χαρακτηριστική καμπύλη ενός ερωτήματος (Item Characteristic Curve-I.C.C.). Σε αυτό το σημείο οφείλουμε, να τονίσουμε πως, υποθέτουμε ότι η ικανότητα ακολουθεί την τυπική κανονική κατανομή ($N(0,1)$) και επομένως άτομα μεσαίας ικανότητας αναμένουμε να πάρουν βαθμολογίες κοντά στο 0, άτομα υψηλής ικανότητας πάνω από 1 και άτομα χαμηλής ικανότητας κάτω από -1.



Εικόνα 1 Η θέση των καμπυλών διαφοροποιείται ανάλογα με τη δυσκολία τους και το επίπεδο ικανότητας των εξετασθέντων

Παρατηρούμε ότι η θέση των καμπυλών οι οποίες αντιστοιχούν στα 3 ερωτήματα διαφοροποιείται ανάλογα με τη δυσκολία τους και το επίπεδο ικανότητας των εξετασθέντων. Οι κλίσεις των καμπυλών οι οποίες συγκλίνουν αλλά δεν διασταυρώνονται είναι ίδιες. Η πιθανότητα απάντησης σε ένα ερώτημα με βάση την οποία διαμορφώνεται η εν λόγω καμπύλη σχετίζεται με το βαθμό δυσκολίας του και τη διακριτική του ισχύ. Η δυσκολία ενός ερωτήματος προσδιορίζει σε ποιο εύρος της κλίμακας ικανότητας αυτό λειτουργεί. Π.χ. ένα εύκολο ερώτημα λειτουργεί μεταξύ ατόμων με χαμηλή ικανότητα θ , ενώ ένα δύσκολο ερώτημα λειτουργεί για άτομα με υψηλή ικανότητα.

Από το παραπάνω σχήμα προκύπτει ότι η πιθανότητα επιτυχίας των ατόμων αυξάνει όσο μεγαλώνει το επίπεδο της λανθάνουσας ικανότητάς τους να απαντήσουν στο σχετικό ερώτη-

μα. Η πιθανότητα αυτή για άτομα πολύ υψηλής ικανότητας είναι σύμφωνα με το εν λόγω σχήμα ίση με το ένα ενώ για τα άτομα πολύ χαμηλής ικανότητας είναι σχεδόν ίση με μηδέν. Η ερώτηση 1 είναι πιο εύκολη μιας και η πιθανότητα σωστής απάντησης είναι πάντα μεγαλύτερη από τις ερωτήσεις 2 και 3 ανεξαρτήτως της ικανότητας του μαθητή. Για παράδειγμα για ένα άτομο που βρίσκεται στη διάμεσο της ικανότητας (πιο ικανό από το 50% των ατόμων) οι πιθανότητες σωστής απάντησης είναι 75%, 50% και 25% για τις ερωτήσεις 1, 2 και 3 αντίστοιχα. Η δυσκολία όμως, απάντησης στα 3 ερωτήματα είναι διαφορετική, όπως φαίνεται από τη θέση που κατέχουν. Το πρώτο είναι το πιο εύκολο και προϋποθέτει μικρότερη ικανότητα απάντησης ενώ το τρίτο είναι το πιο δύσκολο και απαιτεί μεγαλύτερη ικανότητα απάντησης. Αυτό, σημαίνει ότι αν μας είναι η γνωστή η ικανότητα ενός εξεταζόμενου μπορούμε να προβλέψουμε την πιθανότητα απάντησης του χωρίς να του δώσουμε το σχετικό ερώτημα. (Willis, 2013)

Ο δεύτερος δείκτης αφορά στο πόσο καλά ένα ερώτημα μπορεί να διαχωρίζει τους εξεταζόμενους, αυτούς που έχουν ικανότητα κάτω και πάνω από τη θέση που αυτό κατέχει στην κλίμακα ικανότητας. Όσο πιο απότομη γίνεται η κάμψη της καμπύλης στο μέσο επίπεδο ικανότητας τόσο καλύτερη είναι η διακριτικότητα του αντίστοιχου ερωτήματος. Με βάση τους δείκτες αυτούς, μπορούν τα ερωτήματα να χαρακτηρισθούν ως πολύ εύκολα, εύκολα, μέσης δυσκολίας, δύσκολα ενώ από πλευράς διακριτικότητας δύναται να καταταχθούν σε μηδενικής, χαμηλής, μέσης, υψηλής και άριστης διακριτικής ισχύος (Baker, Kim 2004).

2.3.2 Η αναμόρφωση των κανόνων και η διαφοροποίηση

Η νέα θεωρία για τις μετρήσεις έχει διαμορφώσει νέους κανόνες, οι οποίοι διαφέρουν σημαντικά από αυτούς στους οποίους στηρίζεται η κλασική θεωρία. Οι κύριες διαφορές τους συνίσταται, (Κασσωτάκης, 1999) στα εξής

- α)** Το τυπικό σφάλμα μέτρησης δεν είναι το ίδιο για όλους τους εξεταζόμενους που ανήκουν σε ορισμένο πληθυσμό όπως δέχεται η κλασική θεωρία αλλά διαφοροποιείται μεταξύ των επί μέρους σκορ
- β)** τα εκτενή τεστ δεν είναι πάντα πιο αξιόπιστα από τα σύντομα όπως δέχεται η κλασική θεωρία
- γ)** η σύγκριση των αποτελεσμάτων που επιτυγχάνονται με διαφορετικές μορφές τεστ είναι εφικτή ακόμη και αν δεν υπάρχει ισοδύναμη παραλληλία τους, όπως επιβάλλει η κλασική θεωρία
- δ)** η αρχή της κλασικής θεωρίας, κατά την οποία για να γίνουν ακριβείς εκτιμήσεις απαιτείται αντιπροσωπευτικότητα των δειγμάτων στα οποία σταθμίζεται ένα τεστ ανατρέπε-

- ται από τη θεωρία μέτρησης της ικανότητας απαντήσεις σε ερωτήματα αφού δέχεται ότι μπορούν να γίνουν έγκυρες εκτιμήσεις και χωρίς την προϋπόθεση αυτή
- ε) τα αποτελέσματα ενός τεστ μπορούν να ερμηνευθούν όχι με βάση νόρμες αλλά με βάση τη σύγκριση των αποστάσεων τους στην ενιαία κλίμακα ικανότητας
 - στ) όρος ότι οι ιδιότητες των η ισοδιαστημικών κλιμάκων ισχύουν, κατά την κλασική θεωρία όταν υπάρχουν, κανονικές κατανομές των μετρήσεων μεταβάλλεται αφού κατά IRT αυτό ισχύει, όταν εφαρμόζονται τα κατάλληλα μοντέλα μέτρησης
 - ζ) μικτού τύπου ερωτήματα επηρεάζουν, κατά την κλασική θεωρία ανομοιομορφα τα τελικά αποτελέσματα ενώ κατά την IRT αυτό δεν ισχύει.
 - η) Η αλλαγή επιδόσεων μπορεί να συγκριθεί ακόμη και στην περίπτωση ανομοιογενούς αρχικής αφετηρίας, άποψη που δεν γίνεται δεκτή στο πλαίσιο της κλασικής θεωρίας.
 - θ) Η παραγοντική ανάλυση παρέχει πληρέστερη πληροφόρηση στο πλαίσιο IRT από ότι στο πλαίσιο της κλασικής θεωρίας.
 - ι) Τέλος, τα ιδιαίτερα χαρακτηριστικά των ερωτημάτων σχετίζονται με τις ψυχομετρικές τους ιδιότητες, ενώ υποστηρίζεται το αντίθετο από την κλασική θεωρία.

2.4 Τα μοντέλα ανάλυσης

Τα μοντέλα μέτρησης με βάση τα οποία αποτιμάται κατά πόσο οι ιδιότητες των ερωτημάτων (δυσκολία διακριτικότητα, βαθμός πιθανότητας να δοθεί ορθή απάντηση στην τύχη) και το επίπεδο κατοχής του μετρούμενου λανθάνοντος χαρακτηριστικό από ένα άτομο καθορίζουν την πιθανότητα που έχει να απαντήσει επιτυχώς τα ερωτήματα. Τα πρώτα βασίζονται στην υπόθεση ότι η δυνατότητα των εξεταζομένων να απαντήσουν στα ερωτήματα ενός τεστ προσδιορίζεται από ένα κυρίως λανθάνον χαρακτηριστικό σε σχέση με το οποίο έχουν διατυπωθεί τα ερωτήματα που περιλαμβάνονται σε ένα τεστ. Το χαρακτηριστικό αυτό μπορεί, σε ορισμένες περιπτώσεις να έχει σύνθετο χαρακτήρα (π.χ. ένα τεστ μπορεί να μετρά την ικανότητα ανάγνωσης και την ικανότητα κειμένου). Στη δεύτερη περίπτωση δηλαδή πολυδιάστατα μοντέλα, ισχύει η υπόθεση ότι τα μετρούμενα λανθάνοντα χαρακτηριστικά των ατόμων είναι περισσότερα του ενός και εν διαφέρουν μεταξύ τους κατά τρόπο διακριτό. Ένα τεστ π.χ. μπορεί να μετρά την αναγνωστική και τη μαθηματική ικανότητα.

Σε γενικές γραμμές, τα μοντέλα IRT μπορούν να χωριστούν σε δύο οικογένειες: μονοδιάστατα και πολυδιάστατα. Τα μονοδιάστατα μοντέλα απαιτούν μια ιδιότητα μεμονωμένου χαρακτηριστικού (ικανότητας) θ. Πολυδιάστατα μοντέλα IRT μοντελοποιούν δεδομένα απόκρισης που υποτίθεται ότι προκύπτουν από πολλαπλά χαρακτηριστικά. Ωστόσο, λόγω της πολύ αυξημένης πολυπλοκότητας, η πλειοψηφία των ερευνών και εφαρμογών IRT χρησιμο-

ποιεί ένα μονοδιάστατο μοντέλο. Τα μοντέλα IRT (Lord, 1980) μπορούν επίσης να κατηγοριοποιηθούν με βάση τον αριθμό των βαθμολογούμενων απαντήσεων. Το τυπικό στοιχείο πολλαπλής επιλογής είναι διχοτομικό, παρόλο που μπορεί να υπάρχουν τέσσερις ή πέντε επιλογές, εξακολουθεί να βαθμολογείται μόνο ως σωστή/λάθος. Μια άλλη κατηγορία μοντέλων (McCoach, 1980) ισχύει για πολυτομικά αποτελέσματα, όπου κάθε απάντηση έχει διαφορετική τιμή βαθμολογίας. Ένα συνηθισμένο παράδειγμα αυτού είναι τα στοιχεία τύπου κλίμακας Likert, π.χ. "Βαθμολογήστε σε κλίμακα από 1 έως 5". Ακολουθεί ανάλυση των μοντέλων (Baker, Kim 2004),(Baker, 2004) (Lord, 1980) (Thissen, 2001)

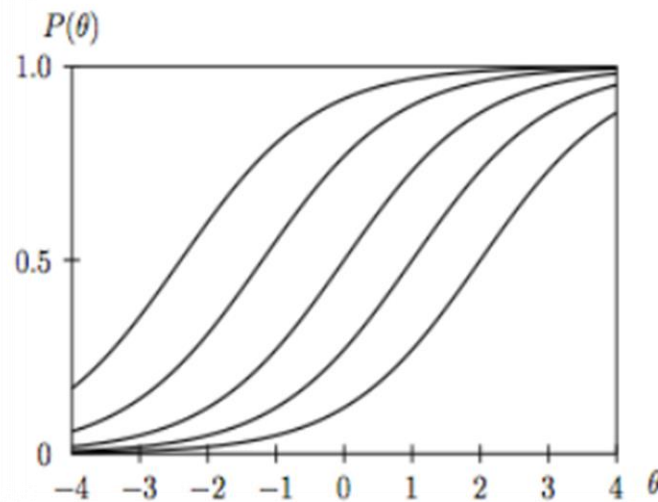
- **Μοντέλο μιας παραμέτρου (1PL-Parameter Logistic Model /RASH)**

Το μοντέλο, αυτό, είναι η απλούστερη μορφή μοντέλων IRT. Αποτελείται από μια παράμετρο που περιγράφει το λανθάνον χαρακτηριστικό (ικανότητα – θ) του ατόμου που ανταποκρίνεται στα ερωτήματα καθώς και από μια άλλη παράμετρο για το αντικείμενο (δυσκολία) (Baker et al., 2004). Η ακόλουθη εξίσωση αντιπροσωπεύει τη μαθηματική του μορφή:

$$P(x_i/\theta, b_i) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

Δηλαδή, Πιθανότητα σωστής απόκρισης στο ερώτημα x_i για συγκεκριμένο λανθάνον χαρακτηριστικό θ , με την παράμετρο δυσκολίας b_i που δείχνει την διαφοροποίηση (μεταξύ των items) ως προς την δυσκολία ή προβλέπει την πιθανότητα μιας σωστής απόκρισης δεδομένης της ικανότητας και της δυσκολίας του ερωτώμενου στο ερώτημα.

Δηλαδή, προβλέπει την πιθανότητα μιας σωστής απόκρισης, δεδομένης της ικανότητας και της δυσκολίας, του ερωτώμενου/υποκειμένου, στο ερώτημα. Στο μοντέλο 1-PL, η παράμετρος διάκρισης είναι σταθερή για όλα τα στοιχεία, και κατά συνέπεια όλες οι Χαρακτηριστικές Καμπύλες Στοιχείων (Item Characteristic Curves) που αντιστοιχούν στα διαφορετικά στοιχεία του μέτρου είναι παράλληλες κατά μήκος της κλίμακας ικανότητας.



Εικόνα 2 Χαρακτηριστικές Καμπύλες Στοιχείων που αντιστοιχούν στα διαφορετικά στοιχεία του μέτρου είναι παράλληλες κατά μήκος της κλίμακας ικανότητας.

Για την κάθε καμπύλη στον κάθετο άξονα αριστερά (η πιθανότητα σωστής απάντησης) και στον κάθετο δεξιά (η αναμενόμενη βαθμολογία) και στον οριζόντιο άξονα (η ικανότητα του ατόμου). Το σχήμα (Εικόνα 2), δείχνει 5 αντικείμενα, το πιο δεξιά είναι το πιο δύσκολο και πιθανότατα θα «αποκριθεί» σωστά από εκείνους με υψηλότερη ικανότητα ενώ το αριστερά υποδηλώνει εύκολο στοιχείο και απαιτεί χαμηλότερα επίπεδα λανθάνοντος χαρακτηριστικού για να επιτευχθεί η ίδια πιθανότητα ορθής απάντησης 50%.

Rasch	$P_{ij}(\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$	θ =ικανότητα b_i =παράμετρος δυσκολίας
1-PL	$P(Y_{is} = 1/\theta_s) = \frac{\exp(1.7a(\theta_s - b_i))}{1 + \exp(1.71a(\theta_s - b_i))}$	θ =ικανότητα b_i = παράμετρος δυσκολίας a =παράμετρος διάκρισης 1.7 =παράγοντας κλίμακας

Τα μοντέλα είναι μαθηματικά ισοδύναμα, ωστόσο, το μοντέλο Rasch περιορίζει το a στο 1, ενώ το μοντέλο 1-Parameter προσπαθεί να προσαρμόσει τα δεδομένα όσο το δυνατόν περισσότερο και δεν περιορίζει τον παράγοντα διάκρισης στο 1. Το πρώτο μοντέλο υπερέχει, κα-

θώς ασχολείται περισσότερο με την ανάπτυξη της μεταβλητής που χρησιμοποιείται για τη μέτρηση της διάστασης ενδιαφέροντος, βελτιώνοντας την ακρίβεια των ερωτημάτων.

Η συνάρτηση πληροφοριών ερωτήματος (Item Information Function) δείχνει τον όγκο των πληροφοριών που παρέχει κάθε στοιχείο και υπολογίζεται πολλαπλασιάζοντας την πιθανότητα να εγκρίνετε μια σωστή απάντηση πολλαπλασιασμένη με την πιθανότητα λανθασμένης απάντησης και δίνεται από τον τύπο.

$I_i(\theta, b_i) = P_i(\theta_j, b_i) Q_i(\theta_j, b_i)$	<p>θ=ικανότητα</p> <p>b_i=παράμετρος δυσκολίας</p> <p>Q=λανθασμένη απάντηση</p> <p>P=σωστή απάντηση</p>
--	--

Πρέπει να σημειωθεί ότι η ποσότητα των πληροφοριών σε ένα δεδομένο επίπεδο ικανότητας είναι το αντίστροφο της διακύμανσής της, επομένως, όσο μεγαλύτερη είναι η ποσότητα των πληροφοριών που παρέχονται από το στοιχείο, τόσο μεγαλύτερη είναι η ακρίβεια της μέτρησης. Καθώς οι πληροφορίες του αντικειμένου σχεδιάζονται έναντι της ικανότητας, ένα αποκαλυπτικό γράφημα απεικονίζει τον όγκο των πληροφοριών που παρέχονται από το αντικείμενο. Τα στοιχεία που μετρούνται με μεγαλύτερη ακρίβεια, παρέχουν περισσότερες πληροφορίες και απεικονίζονται γραφικά ως μεγαλύτερα και στενότερα, σε σύγκριση με τα αντίστοιχα που παρέχουν λιγότερες πληροφορίες. Η κορυφή της καμπύλης αντιστοιχεί με την τιμή του b_i – η ικανότητα στο σημείο της διάμεσης πιθανότητας. Ο μέγιστος αριθμός παρεχόμενων πληροφοριών θα δίνεται όταν η πιθανότητα να απαντήσετε σωστά ή λάθος είναι ίση, δηλαδή 50%. Τα στοιχεία είναι πιο ενημερωτικά μεταξύ των ερωτηθέντων που αντιπροσωπεύουν ολόκληρη τη λανθάνουσα συνέχεια και ειδικά μεταξύ εκείνων που έχουν 50% πιθανότητα να απαντήσουν με κάθε τρόπο.

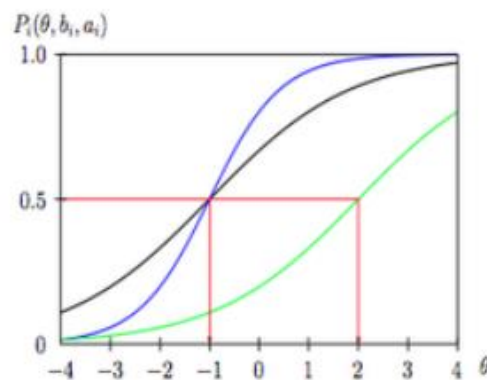
- **Μοντέλο Δύο Παραμέτρων (2-PL)**

Το μοντέλο δύο παραμέτρων προβλέπει την πιθανότητα μιας επιτυχούς απάντησης χρησιμοποιώντας δύο παραμέτρους b_i και a_i .

$P_{ij} = (\theta_j, b_i, a_i) = \frac{\exp[a_i(\theta_j - b_i)]}{1 + \exp[a_i(\theta_j - b_i)]}$	<p>θ=ικανότητα</p> <p>b_i=παράμετρος δυσκολίας</p> <p>a_i=παράμετρος διάκρισης η οποία</p>
---	--

μπορεί να διαφέρει από Item σε Item

Η παράμετρος διάκρισης επιτρέπεται να ποικίλλει μεταξύ των στοιχείων. Στο εξής, το ICC των διαφορετικών ερωτημάτων (Εικόνα 2) μπορεί να τέμνεται και να έχει διαφορετικές κλίσεις. Όσο πιο απότομη είναι η κλίση, τόσο μεγαλύτερη είναι η διάκριση του αντικειμένου, καθώς θα μπορεί να ανιχνεύσει ανεπαίσθητες διαφορές στην ικανότητα των ερωτηθέντων.



Εικόνα 3 Χαρακτηριστικές Καμπύλες που τέμνονται

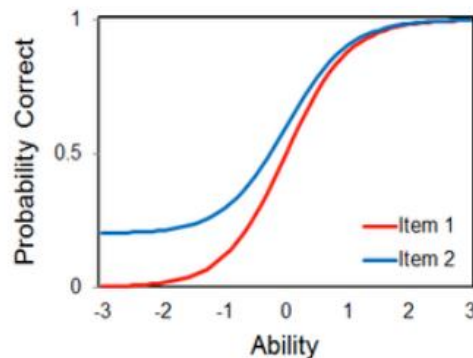
Όπως συμβαίνει με το μοντέλο μιας παραμέτρου, οι πληροφορίες υπολογίζονται ως το γινόμενο μεταξύ της πιθανότητας μιας σωστής και μιας λανθασμένης απάντησης. Ωστόσο, το γινόμενο πολλαπλασιάζεται με το τετράγωνο της παραμέτρου διάκρισης. Το συμπέρασμα είναι ότι, όσο μεγαλύτερη είναι η παράμετρος διάκρισης, τόσο μεγαλύτερη είναι η πληροφόρηση που παρέχει το αντικείμενο. Καθώς ο παράγοντας διάκρισης επιτρέπεται να ποικίλλει μεταξύ των στοιχείων, τα γραφήματα συναρτήσεων πληροφοριών στοιχείων μπορεί επίσης να φαίνονται διαφορετικά.

$$I_i(\theta, b_i, a_i) = a_i^2 P_i(\theta, b_i) Q_i(\theta, b_i)$$

- **Μοντέλο Τριών Παραμέτρων (3-PL)**

Το μοντέλο προβλέπει την πιθανότητα μιας σωστής απάντησης, με τον ίδιο τρόπο όπως το μοντέλο 1 – PL και το μοντέλο 2 PL – αλλά περιορίζεται από μια τρίτη παράμετρο που ονομάζεται παράμετρος εικασίας (επίσης γνωστή ως παράμετρος ψευδοτυχίας), η οποία πε-

ριορίζει η πιθανότητα να επικυρωθεί μια σωστή απάντηση όταν η ικανότητα του ερωτώμενου πλησιάζει $-\infty$.



Εικόνα 4 Σιγμοειδείς Καμπύλες

Καθώς οι ερωτώμενοι απαντούν σε ένα στοιχείο μαντεύοντας, ο όγκος των πληροφοριών που παρέχονται από αυτό το στοιχείο μειώνεται και η συνάρτηση του στοιχείου πληροφοριών κορυφώνεται σε χαμηλότερο επίπεδο σε σύγκριση με άλλες συναρτήσεις. Επιπλέον, η δυσκολία δεν ορίζεται πλέον στη διάμεση πιθανότητα. Τα ερωτήματα που απαντώνται με εικασία υποδεικνύουν ότι η ικανότητα του ερωτώμενου είναι μικρότερη από τη δυσκολία του.

$$P_i = (\theta, \alpha, b, c) = c_i + (1 - c_i) \frac{\exp[\alpha_i(\theta_j - b_i)]}{1 + \exp[\alpha_i(\theta_j - b_i)]}$$

θ =ικανότητα

b_i =δυσκολία παραμέτρου

α_i =παράμετρος διάκρισης η οποία μπορεί να διαφέρει από Item σε Item

c =παράμετρος πρόβλεψης-εικασίας

Το φλέγον ζήτημα στις αναλύσεις των δεδομένων είναι το πως επιλέγουμε το κατάλληλο μοντέλο κάθε φορά. Ένας τρόπος για την επιλογή του καταλληλότερου μοντέλου είναι να αξιολογήσουμε τη σχετική προσαρμογή του μοντέλου μέσω των κριτηρίων πληροφοριών του. Τα κριτήρια BIC (Bayesian Information Criterion) και AIC (Akaike's Information Criteria) είναι ελαφρώς διαφορετικά, αλλά κανένα δεν είναι καλύτερο από το άλλο οπότε ανεξάρτητα με το πιο θα επιλέξουμε θα επιλέξουμε με βάση την χαμηλότερη τιμή για να αποφανθούμε. Καθώς η πολυπλοκότητα του μοντέλου αυξάνεται, η τιμή AIC/BIC αυξάνεται και όσο αυξάνεται η πιθανότητα, η AIC/BIC μειώνεται. Άρα, το χαμηλότερο είναι καλύτερο. Επιπλέον, με το χ^2 -test πραγματοποιούμε έλεγχο για την καλή προσαρμογή των ερωτήσεων

στο επιλεχθέν μοντέλο. Τα άνωθεν, θα παρουσιαστούν αναλυτικά στο επόμενο Κεφάλαιο 4 όπου και θα βρουν εφαρμογή.

2.5 Κριτική Ανασκόπηση

Πολλές μελέτες έχουν διεξαχθεί διερευνώντας τη σχέση και τις διαφορές των CCT και IRT. Παρουσιάζονται κάποιες έρευνες, τυχαία, οι οποίες εργάστηκαν και με τις δύο μεθόδους κι έκλιναν προς την IRT.

Ojerinde et al. (Ruslan, 2016) αξιολογούν τη χρήση δεδομένων από pretest αγγλικών έτσι ώστε να συγκριθούν οι δείκτες που ελήφθησαν χρησιμοποιώντας το μοντέλο 3 παραμέτρων της Θεωρίας Απόκρισης Αντικειμένων (IRT) και την προσέγγιση της Κλασικής Θεωρίας Δοκιμών Classical Test Theory (CTT) και εξακριβώνουν πόσο καλά μπορούν και οι δύο να προβλέψουν τα πραγματικά αποτελέσματα των δοκιμών και το βαθμό τους, τη συγκρισιμότητα, χρησιμοποιώντας ένα δείγμα 1075 εξεταζομένων που "πήρε" μια έκδοση του pre-test. Τα ευρήματα αυτής της μελέτης έχουν δείξει ότι τα στατιστικά στοιχεία ατόμου και αντικειμένου που προέρχονται από το δύο πλαίσια μέτρησης είναι αρκετά συγκρίσιμα. Ο βαθμός της εσωτερικής διασποράς των στατιστικών στοιχείων κατά μήκος του δείγματος, που συνήθως θεωρούνται ως θεωρητικά ανωτερότητα των μοντέλων IRT, φάνηκε επίσης να είναι παρόμοια για τα δύο πλαίσια μέτρησης αλλά το μοντέλο IRT έδωσε μια καλύτερη ιδέα για το εσωτερική συνοχή της δοκιμής από το CTT. Επιπλέον, το 3PL μοντέλο βρέθηκε να είναι πιο κατάλληλο σε ερωτήσεις πολλαπλής επιλογής αν και περιλάμβανε λίγο πιο σύνθετη διαδικασία μαθηματικής εκτίμησης από το CTT. Abedalaziz and Leng (Ruslan, 2016) στη μελέτη για τη σχέση μεταξύ CTT και IRT Approaches in Analysing Item Characteristics, ο στόχος ήταν να συγκριθεί η δυσκολία και το στοιχείο διάκρισης Μαθηματικής ικανότητας χρησιμοποιώντας τις δύο μεθόδους σε 1, 2 και 3 παραμέτρους. Τα δεδομένα που συγκεντρώθηκαν αναλύθηκαν με τα αποτελέσματα να δείχνουν ότι το λογιστικό μοντέλο 3 παραμέτρων έχει τους περισσότερους συγκρίσιμους δείκτες με CTT, επιπλέον, CTT και Μοντέλα IRT μπορούν να χρησιμοποιηθούν συνεργατικά. Ojerinde (2013) διεξήγαγε μια μελέτη για την αξιολόγηση της χρησιμότητας των ψυχομετρικών δεδομένων που λαμβάνονται με τη χρήση και των δύο μοντέλων στην ανάλυση του UTME Physics Pre-test ώστε να εξεταστούν τα αποτελέσματα που προέκυψαν και να καθοριστούν πόσο καλά μπορούν να προβλέψουν τα πραγματικά αποτελέσματα των δοκιμών και το βαθμός συγκρισιμότητας τους. Ο ερευνητής επίσης επαλήθευσε τις προϋποθέσεις που πρέπει να πληρούνται ώστε το IRT να εφαρμόζεται με πραγματικά δεδομένα δοκιμών. Τα ευρήματα έδειξαν ότι το αποτέλεσμα που προέκυψε χρησιμοποιώντας το μοντέλο IRT ήταν ότι είναι πιο κατάλληλο σε ερωτήσεις πολλαπλής επιλογής (MCQs) στο τεστ ικανότητας αλλά περιλάμβανε πιο σύνθετη διαδικασία μαθηματικής εκτίμησης από την κλασική προσέγγιση. Συνολικά, προέκυψαν δείκτες και από τις δύο προσεγγίσεις έδωσε πολύτιμες πληροφορίες με συγκρίσιμα και σχεδόν εναλλάξιμα αποτελέσματα σε ορισμένες περιπτώσεις. Σε επιπλέον, ακόμα μία μελέτη, ο Guler et al. (Ruslan, 2016) σύγκρινε την κλασική θεωρία δοκιμών και θεωρία απόκρισης στοιχείων σε όρους των παραμέτρων των στοιχείων. Στόχος της μελέτης τους ήταν να εξετάσει εμπειρικά τις ομοιότητες και διαφορές

στις παραμέτρους που υπολογίζονται με χρήση των δύο προσεγγίσεων, ένα τυχαίο δείγμα μαθητών που συμμετείχαν στις Εισαγωγικές εξετάσεις τουρκικών λυκείων (HSEE). Τα ευρήματα αποκαλύπτουν ότι, τις υψηλότερες συσχετίσεις μεταξύ CTT και μιας παραμέτρου IRT (0,99) ως προς τη δυσκολία του είδους παραμέτρων και μεταξύ CTT και IRT 2 παραμέτρων μοντέλο (0,96) όσον αφορά στη παράμετρο διάκρισης. Το αποτέλεσμα δείχνει επίσης χαμηλότερο επίπεδο συσχέτιση μεταξύ του μοντέλου 3 παραμέτρων και του CTT, αν και το μοντέλο 3 παραμέτρων προσδιορίστηκε ως το πιο ταιριαστό από την άποψη της προσαρμογής μοντέλου-δεδομένων. Υπό το φως των ευρημάτων τους, μπορεί να λεχθεί ότι δεν υπάρχει μεγάλη διαφορά μεταξύ της χρήσης IRT 1 ή 2 παραμέτρων μοντέλο και CTT. Ωστόσο, σε περιπτώσεις όπου η πιθανότητα εικασίας είναι υψηλή, υπάρχει σημαντική διαφορά μεταξύ μοντέλου 3 παραμέτρων και CTT.

Η CTT και η IRT ασχολούνται σε μεγάλο βαθμό με τα ίδια προβλήματα, αλλά είναι διαφορετικά σώματα θεωρίας και συνεπώς διαφορετικές μεθόδους. Και οι δύο θεωρίες επιτρέπουν την πρόβλεψη των αποτελεσμάτων των ψυχολογικών τεστ προσδιορίζοντας τις παραμέτρους της δυσκολίας των αντικειμένων και την ικανότητα των υποψηφίων. Και οι δύο ενδιαφέρονται να βελτιώσουν την αξιοπιστία και την εγκυρότητα των ψυχολογικών τεστ. Και οι δύο αυτές προσεγγίσεις παρέχουν μέτρα εγκυρότητας και αξιοπιστίας. Υπάρχουν ορισμένα εντοπισμένα ζητήματα στην κλασική θεωρία δοκιμών που αφορούν τη βαθμονόμηση της δυσκολίας του αντικειμένου, την εξάρτηση του δείγματος από τα μέτρα συντελεστών και τις εκτιμήσεις του σφάλματος μέτρησης, τα οποία με τη σειρά τους αντιμετωπίζονται από τη θεωρία απόκρισης στοιχείων.

Αν και τα δύο είναι γενικά λειτουργούν καλά και συμπληρωματικά, υπάρχουν ορισμένα σημεία διαφοράς:

- Η IRT κάνει ισχυρότερες υποθέσεις από το CTT και σε πολλές περιπτώσεις παρέχει αντίστοιχα ισχυρότερα ευρήματα. Φυσικά, αυτά τα αποτελέσματα ισχύουν μόνο όταν πληρούνται πραγματικά οι υποθέσεις των μοντέλων IRT.
- Αν και τα αποτελέσματα του CTT επέτρεψαν σημαντικά πρακτικά αποτελέσματα, η βασισμένη-σε μοντέλα-φύση του IRT προσφέρει πολλά πλεονεκτήματα σε σχέση με τα ανάλογα ευρήματα CTT.
- Οι διαδικασίες βαθμολόγησης δοκιμών CTT έχουν το πλεονέκτημα ότι είναι απλές στον υπολογισμό (και στην εξήγηση) ενώ η βαθμολογία IRT απαιτεί γενικά σχετικά περίπλοκες διαδικασίες εκτίμησης.
- Το IRT παρέχει αρκετές βελτιώσεις στην κλιμάκωση στοιχείων και ατόμων. Οι ιδιαιτερότητες εξαρτώνται από το μοντέλο IRT, αλλά τα περισσότερα μοντέλα κλιμακώνουν τη δυσκολία των αντικειμένων και την ικανότητα των ατόμων στην ίδια μέτρηση. Έτσι, η δυσκολία ενός αντικειμένου και η ικανότητα ενός ατόμου μπορούν να συγκριθούν με νόημα.
- Μια άλλη βελτίωση που παρέχεται από το IRT είναι ότι οι παράμετροι των μοντέλων IRT γενικά δεν εξαρτώνται από το δείγμα ή τη δοκιμή, ενώ η πραγματική βαθμολογία ορίζεται στο CTT στο πλαίσιο μιας συγκεκριμένης δοκιμής. Έτσι το IRT παρέχει σημαντικά μεγαλύτερη ευελιξία σε καταστάσεις όπου χρησιμοποιού-

νται διαφορετικά δείγματα ή φόρμες δοκιμής. Αυτά τα ευρήματα IRT είναι θεμελιώδη για τις προσαρμοστικές δοκιμές μέσω υπολογιστή.

Πίνακας 1 Σύγκριση Μεθόδων

Χαρακτηριστικό	CTT	IRT
Σύνθεση	Βασίζεται σε απλά μαθηματικά και στην ουσία πρόκειται για τη μέθοδο που χρησιμοποιείται στη βαθμολόγηση μαθητών/φοιτητών στην Ελλάδα.	Σύνθετη προσέγγιση για την ανάλυση αλλά δεν είναι μόνο για ανάλυση. Είναι ένα πλήρες ψυχομετρικό παράδειγμα που αλλάζει τον τρόπο με τον οποίο αναπτύσσονται οι τράπεζες θεμάτων και παράγονται βαθμολογίες. Υπάρχουν πολλά οφέλη σε αυτή την προσέγγιση που δικαιολογούν την πολυπλοκότητα και υπάρχει λόγος που όλες οι μεγάλες εξετάσεις στον κόσμο χρησιμοποιούν το IRT.
Στατιστικά	Τα στατιστικά CTT για τις συνολικές βαθμολογίες και τις δευτερεύουσες βαθμολογίες περιλαμβάνουν αξιοπιστία συντελεστή άλφα, τυπικό σφάλμα μέτρησης (συνάρτηση αξιοπιστίας και SD), περιγραφικά στατιστικά στοιχεία (μέσος όρος, SD...) και συνοψίσεις στατιστικών στοιχείων (π.χ. μέσος όρος Rpbis).	Με το IRT, χρησιμοποιούμε τα ίδια περιγραφικά στατιστικά στοιχεία, αλλά οι βαθμολογίες είναι πλέον διαφορετικές (θήτα, όχι σωστός αριθμός). Το τυπικό σφάλμα μέτρησης είναι πλέον συνάρτηση υπό όρους, όχι ένας μεμονωμένος αριθμός. Ολόκληρη η έννοια της αξιοπιστίας εγκαταλείπεται και αντικαθίσταται με την έννοια της ακρίβειας, και επίσης ως η ίδια συνάρτηση υπό όρους.
Ανάλυση σε επίπεδο ερωτήματος	Τα στατιστικά στοιχεία για το CTT περιλαμβάνουν την ορθή αναλογία (δυσκολία), τη συσχέτιση σημείων (Rpbis) (διάκριση) και μια ανάλυση διασπαστών/απάντησης. Εάν υπάρχουν δημογραφικές πληροφορίες, η ανάλυση CTT μπορεί επίσης να παρέχει μια απλή αξιολόγηση της διαφορικής λειτουργίας στοιχείων (DIF).	Το IRT θεωρεί ότι η πιθανότητα σωστής απάντησης εξαρτάται από το επίπεδο δυσκολίας της ερώτησης, την ικανότητα του μαθητή, την περίπτωση τυχαίας απάντησης και της ικανότητας της ερώτησης να ξεχωρίζει μαθητές διαφορετικής ικανότητας. Το πιο σημαντικό είναι ότι δημιουργεί εντελώς νέες κατηγορίες στατιστικών για στοιχεία κλίμακας αξιολόγησης
Ως προς τη	Το CTT βαθμολογεί τα τεστ με την παραδοσιακή βαθμολόγηση:	Το IRT βαθμολογεί τους εξεταζόμενους απευθείας σε μια λανθά-

βαθμολογία	αριθμός-σωστός, ορθός αναλογίας ή άθροισμα βαθμών.	νουςα κλίμακα.
Σύνδεση και εξισορρόπηση	Η σύνδεση και η εξισορρόπηση είναι μια στατιστική ανάλυση για τον προσδιορισμό συγκρίσιμων βαθμολογιών σε διαφορετικές μορφές. π.χ., το Έντυπο Α είναι «δύο σημεία ευκολότερο» από το Έντυπο Β και επομένως το 72 στη Μορφή Α είναι συγκρίσιμο με το 70 στη Μορφή Β. Το CTT έχει διάφορες μεθόδους για αυτό, συμπεριλαμβανομένων των μεθόδων Tucker και Levine, αλλά υπάρχουν μεθοδολογικά προβλήματα με αυτές προσεγγίσεις. Αυτά τα ζητήματα, και άλλα θέματα με το CTT, οδήγησαν τελικά στην ανάπτυξη του IRT στις δεκαετίες του 1960 και του 1970.	Το IRT έχει μεθόδους για την επίτευξη σύνδεσης και εξισορρόπησης που είναι πολύ πιο ισχυρές από το CTT, συμπεριλαμβανομένων μεθόδων βαθμονόμησης στοιχείων ή μετατροπής όπως το Stocking-Lord.
Μέγεθος Δείγματος	Η κλασική θεωρία δοκιμών μπορεί να λειτουργήσει αποτελεσματικά με 50 εξεταζόμενους και να παρέχει χρήσιμα αποτελέσματα με μόλις 20.	Ανάλογα με το μοντέλο IRT που επιλέγεται το ελάχιστο μέγεθος δείγματος μπορεί να φτάσει έως 1.000.
Σχεδιασμός και Κατασκευή φόρμας Τεστ	Η Κλασική Θεωρία Δοκιμών και η Θεωρία Απόκρισης Ερωτημάτων διαφέρουν ως προς τον τρόπο σχεδιασμού και κατασκευής των φορμών δοκιμής. Η κλασική θεωρία των τεστ λειτουργεί καλύτερα όταν υπάρχουν πολλά στοιχεία μέσης δυσκολίας, καθώς αυτό μεγιστοποιεί την αξιοπιστία του συντελεστή άλφα.	Υπάρχουν σίγουρα περιπτώσεις όπου ο σκοπός της αξιολόγησης είναι διαφορετικός. Το IRT παρέχει ισχυρότερες μεθόδους για το σχεδιασμό τέτοιων δοκιμών και στη συνέχεια τη βαθμολόγηση.
Παράμετρος Εικασίας	Η κλασική θεωρία δοκιμών δεν έχει αποτελεσματικό τρόπο να εξηγήσει την εικασία.	Η θεωρία απόκρισης αντικειμένων έχει μια παράμετρο για να εξηγήσει την εικασία.
Παράμετρος Δυσκολίας και Διάκρισης	δυσκολία και διάκριση – εξαρτώνται από την ομάδα στην οποία εκτιμώνται και η αμερόληπτη εκτίμηση αυτών των παραμέτρων απαιτεί αντιπροσωπευ-	Δυσκολία και διάκριση – που υπολογίζονται σε ένα δείγμα από έναν πληθυσμό μπορούν γραμμικά να μετασχηματιστούν σε εκτιμήσεις αυτών των παραμέτρων σε άλλο

	τικό δείγμα.	δείγμα από τον ίδιο πληθυσμό. Οι αμερόληπτες εκτιμήσεις δεν απαιτούν αντιπροσωπευτικό δείγμα.
Εξαρτήσεις	Εξαρτάται από την ομάδα (δηλαδή, βασίζεται στην τυπική απόκλιση της ομάδας) και είναι σταθερό για μια ομάδα, ανεξάρτητα από το επίπεδο βαθμολογίας.	Ανεξάρτητο από την ομάδα στις μετρήσεις που λαμβάνονται. Το IRT επιτρέπει τον υπολογισμό ενός SEM για ένα μεμονωμένο άτομο με βάση την απόδοσή του και τις παραμέτρους του στοιχείου. Παρέχει επίσης SEM υπό όρους που επιτρέπουν στο SEM να ποικίλλει στα διαφορετικά επίπεδα λανθάνοντος συνεχούς.
Προσέγγιση	Γραμμικό	Μη γραμμικό
Έκταση φόρμας δοκιμής	Οι μεγαλύτερες δοκιμές (τεστ) είναι πιο αξιόπιστες από τις μικρότερες δοκιμές.	Τα πιο σύντομα τεστ μπορεί να είναι πιο αξιόπιστα από τα μεγαλύτερα τεστ.

Αξίζει, να σημειωθεί ότι η εκτίμηση του αξιολογούμενου χαρακτηριστικού βασίζεται, σύμφωνα με την κλασική θεωρία, στο άθροισμα του τελικού αριθμού των ορθών απαντήσεων σε ένα τεστ και στην αναγωγή του όπου απαιτείται σε πρότυπους βαθμούς. Αντίθετα, η εκτίμηση αυτή είναι κατά την IRT περισσότερο μια διαδικασία αναζήτησης του ενός εξεταζομένου παρά ένα απλό άθροισμα επιτυχών απαντήσεων. Για το λόγο αυτό τα ερωτήματα ενός τεστ ενδέχεται να μην έχουν ίση βαρύτητα στην τελική εκτίμηση της ικανότητας αυτής, ανάλογα με το μοντέλο που εφαρμόζεται. Το IRT είναι πολύ ισχυρό και μπορεί να παρέχει πρόσθετες πληροφορίες σχετικά με τις δοκιμές εάν χρησιμοποιείται μόνο για την ανάλυση των αποτελεσμάτων για την αξιολόγηση της απόδοσης των στοιχείων και των δοκιμών. Ωστόσο, το IRT είναι πραγματικά χρήσιμο μόνο εάν πρόκειται να το κάνετε το ψυχομετρικό σας παράδειγμα, χρησιμοποιώντας το στη λίστα των παραπάνω δραστηριοτήτων, ειδικά στη βαθμολογία IRT των εξετάσεων. Διαφορετικά, η ανάλυση IRT είναι απλώς ένας άλλος τρόπος εμφάνισης της απόδοσης δοκιμής και αντικειμένου που θα συσχετιστεί ουσιαστικά με το CTT.

ΚΕΦΑΛΑΙΟ 3

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

3.1 Δόμηση ενός Ερωτηματολογίου

3.2 Ερωτηματολόγιο Μαθητών

3.3 Ερωτηματολόγιο Φοιτητών

Το μεγαλύτερο μέρος της βιβλιογραφίας των μεθόδων έρευνας που σχετίζεται με τη χρήση ερωτηματολογίων επικεντρώνεται σε ζητήματα που σχετίζονται με την αξιοπιστία και την εγκυρότητα. Το πρωταρχικό μέλημα είναι οι τρόποι σχεδιασμού ερωτηματολογίων έτσι ώστε να μεγιστοποιηθεί η ακρίβεια και να μειωθεί η μεροληψία στα δεδομένα που συλλέγονται. Συνιστάται προσοχή στο μήκος, τη διάταξη, την αναγνωσιμότητα, τη γλώσσα που χρησιμοποιείται, τη σειρά των ερωτήσεων και το περιεχόμενο των ερωτηματολογίων προκειμένου να βελτιωθεί η ποιότητα των δεδομένων που λαμβάνονται (Thissen, 2001).

3.1 Δόμηση ενός ερωτηματολογίου

Για τον σχεδιασμό ερωτηματολογίων με ποιότητα που δεν στερούνται περιεχομένου και ικανοποιεί τα όσα προαναφέρθηκαν υπάρχει μια συστηματική διαδικασία επτά βημάτων για τον υψηλό σχεδιασμό τους.

- **Βήμα 1: Διεξαγωγή βιβλιογραφικής έρευνας**

Το πρώτο βήμα για την ανάπτυξη ενός ερωτηματολογίου είναι η διεξαγωγή βιβλιογραφικής ανασκόπησης. Υπάρχουν δύο πρωταρχικοί στόχοι για την ανασκόπηση της βιβλιογραφίας: να οριστεί με σαφήνεια η κατασκευή και να προσδιοριστεί εάν υπάρχουν ήδη μέτρα της κατασκευής (ή σχετικών κατασκευών). Η ανασκόπηση της βιβλιογραφίας βοηθά να διασφαλιστεί ότι ο ορισμός της κατασκευής ευθυγραμμίζεται με τη σχετική θεωρία και την έρευνα στο πεδίο, ενώ ταυτόχρονα βοηθά τον ερευνητή να εντοπίσει κλίμακες έρευνας ή στοιχεία

που θα μπορούσαν να χρησιμοποιηθούν ή να προσαρμοστούν για τον τρέχοντα σκοπό (Gehlbach, 2012).

- **Βήμα 2: Διεξαγωγή συνεντεύξεων ή/και ομάδων εστίασης**

Μόλις η ανασκόπηση της βιβλιογραφίας δείξει ότι είναι απαραίτητο να αναπτυχθεί ένα νέο ερωτηματολόγιο και βοήθησε στον ορισμό της κατασκευής, το επόμενο βήμα είναι να εξακριβωθεί εάν η σύλληψη της δομής ταιριάζει με τον τρόπο με τον οποίο οι υποψήφιοι ερωτηθέντες σκέφτονται γι' αυτό (Gehlbach, 2012). Με άλλα λόγια, οι ερωτηθέντες περιλαμβάνουν και αποκλείουν τα ίδια χαρακτηριστικά της δομής με αυτά που περιγράφονται στη βιβλιογραφία; Ποια γλώσσα χρησιμοποιούν οι ερωτηθέντες όταν περιγράφουν την κατασκευή; Για να απαντηθούν αυτά τα ερωτήματα και να διασφαλιστεί ότι το κατασκεύασμα ορίζεται από πολλαπλές προοπτικές, οι ερευνητές θα θέλουν συνήθως να συλλέγουν δεδομένα απευθείας από άτομα που μοιάζουν πολύ με τον πληθυσμό που τους ενδιαφέρει. Το τελικό αποτέλεσμα αυτών των συνεντεύξεων και/ή των ομάδων εστίασης θα πρέπει να είναι μια λεπτομερής περιγραφή του τρόπου με τον οποίο οι πιθανοί ερωτηθέντες αντιλαμβάνονται και κατανοούν την κατασκευή. Αυτά τα δεδομένα θα χρησιμοποιηθούν στη συνέχεια στα Βήματα 3 και 4.

- **Βήμα 3: Σύνθεση κοινού λεξιλογίου**

Αρκετές φορές, υπάρχουν δεδομένα που είναι παρόμοια εννοιολογικά, αλλά η βιβλιογραφία και οι πιθανοί ερωτηθέντες περιγράφουν την κατασκευή χρησιμοποιώντας διαφορετική ορολογία, συνεπώς, είναι λογικό να δομηθεί το ερωτηματολόγιο με «κοινό» λεξιλόγιο, αυτό των πιθανών ερωτηθέντων. Για παράδειγμα, (Gehlbach, 2012) κατά την αξιολόγηση της εμπιστοσύνης του δασκάλου (μερικές φορές αναφέρεται ως αυτο-αποτελεσματικότητα του δασκάλου), είναι πιθανώς πιο κατάλληλο να ρωτήσετε τους δασκάλους σχετικά με την «εμπιστοσύνη τους στη δοκιμή νέων τεχνικών διδασκαλίας» παρά να τους ρωτήσετε για την «αποτελεσματικότητά τους στον πειραματισμό με νέες παιδαγωγικές».

- **Βήμα 4: Ανάπτυξη**

Ένα σημαντικό στοιχείο σχεδιασμού είναι ο αριθμός των ερωτημάτων που απαιτούνται για την επαρκή αξιολόγηση. Δεν υπάρχει εύκολη απάντηση σε αυτή την ερώτηση. Το ιδανικό πλήθος ερωτημάτων, εξαρτάται από διάφορους παράγοντες, συμπεριλαμβανομένης της πολυπλοκότητας της κατασκευής του. Γενικά, είναι καλή πρακτική η ανάπτυξη περισσότερων στοιχείων από αυτά που θα χρειαστούν τελικά στην τελική κλίμακα (π.χ. ανάπτυξη 15 πιθανών στοιχείων με την ελπίδα να δημιουργηθεί τελικά μια κλίμακα οκτώ στοιχείων), επειδή

ορισμένα στοιχεία πιθανότατα θα διαγραφούν ή θα αναθεωρηθούν αργότερα στο η διαδικασία σχεδιασμού (Gehlbach, 2012). Τελικά, η απόφαση για τον αριθμό των αντικειμένων είναι θέμα υποκειμενικής κρίσης. Μόλις οι σχεδιαστές της έρευνας ολοκληρώσουν τη σύνταξη των στοιχείων τους και την επιλογή των ακυρώσεων των απαντήσεών τους, υπάρχουν διάφορες πηγές αποδεικτικών στοιχείων που θα μπορούσαν να χρησιμοποιηθούν για την αξιολόγηση της εγκυρότητας του ερωτηματολογίου και της χρήσης για την οποία προορίζεται. Αυτές οι πηγές εγκυρότητας έχουν περιγράψει στα Πρότυπα για Εκπαιδευτικές και Ψυχολογικές Δοκιμές (Bichi et al., 2015) ως στοιχεία που βασίζονται στα ακόλουθα:

- α) περιεχόμενο
- β) διαδικασία απόκρισης
- γ) εσωτερική δομή, σχέσεις με άλλες μεταβλητές και
- δ) συνέπειες

Τα επόμενα τρία βήματα της διαδικασίας σχεδιασμού ταιριάζουν όμορφα σε αυτήν την ταξινόμηση και περιγράφονται παρακάτω.

▪ **Βήμα 5: Επικύρωση**

Μόλις καθοριστεί η κατασκευή και γραφτούν τα προσχέδια, ένα σημαντικό βήμα στην ανάπτυξη ενός νέου ερωτηματολογίου είναι να αρχίσει η συλλογή αποδεικτικών στοιχείων εγκυρότητας με βάση το περιεχόμενο της έρευνας (η λεγόμενη εγκυρότητα περιεχομένου) (Bichi et al., 2015) . Αυτό το βήμα περιλαμβάνει τη συλλογή δεδομένων από ειδικούς στο περιεχόμενο για να διαπιστωθεί ότι τα μεμονωμένα στοιχεία της έρευνας σχετίζονται με το κατασκεύασμα που μετράται και ότι τα βασικά στοιχεία ή δείκτες δεν έχουν παραλειφθεί (Polit, 2006). Η χρήση ειδικών για τη συστηματική αναθεώρηση του περιεχομένου της έρευνας μπορεί να βελτιώσει σημαντικά τη συνολική ποιότητα και την αντιπροσωπευτικότητα των στοιχείων της κλίμακας (Ruslan, 2016). Οι βασικοί τομείς, που πρέπει να αξιολογηθούν μέσω μιας διαδικασίας επικύρωσης από εμπειρογνώμονες είναι η αντιπροσωπευτικότητα, η σαφήνεια, η συνάφεια και η κατανομή. Η αντιπροσωπευτικότητα ορίζεται ως το πόσο πλήρως τα στοιχεία (στο σύνολό τους) περικλείουν την κατασκευή, η σαφήνεια είναι το πόσο καθαρά διατυπώνονται τα στοιχεία και η συνάφεια αναφέρεται στον βαθμό που κάθε στοιχείο σχετίζεται πραγματικά με συγκεκριμένες πτυχές της κατασκευής. Η κατανομή ενός στοιχείου δεν μετριέται πάντα κατά την επικύρωση από εμπειρογνώμονες, καθώς αναφέρεται στην πιο λεπτή πτυχή του πόσο «δύσκολο» θα ήταν για έναν ερωτώμενο να επιλέξει υψηλή βαθμολογία σε ένα συγκεκριμένο αντικείμενο.

▪ **Βήμα 6: Διεξαγωγή γνωστικών συνεντεύξεων**

Αφού οι ειδικοί βοηθήσουν στη βελτίωση των στοιχείων της κλίμακας, είναι σημαντικό να συλλέξετε αποδεικτικά στοιχεία για την εγκυρότητα της διαδικασίας απόκρισης για να αξιολογήσετε τον τρόπο με τον οποίο οι υποψήφιοι συμμετέχοντες ερμηνεύουν τα στοιχεία και τις απόκρισης (Bichi et al., 2015) . Ένα μέσο συλλογής τέτοιων στοιχείων επιτυγχάνεται μέσω μιας διαδικασίας που είναι γνωστή ως γνωστική συνέντευξη ή γνωστική προ-δοκιμή (Willis, 2013). Παρόμοια με τον τρόπο με τον οποίο οι ειδικοί χρησιμοποιούνται για τον προσδιορισμό της εγκυρότητας περιεχομένου μιας νέας έρευνας, είναι εξίσου σημαντικό να προσδιοριστεί ο τρόπος με τον οποίο οι πιθανοί ερωτηθέντες ερμηνεύουν τα στοιχεία και εάν η ερμηνεία τους ταιριάζει με αυτό που έχει κατά νου ο σχεδιαστής της έρευνας (Willis, 2013). Τα αποτελέσματα από τις γνωστικές συνεντεύξεις μπορεί να είναι χρήσιμα στον εντοπισμό λαθών των ερωτηθέντων κάνουν στην ερμηνεία τους για το στοιχείο ή τις επιλογές απάντησης. Επιπλέον, η αναγκαιότητα για τους ερωτηθέντες να κατανοήσουν κάθε στοιχείο με τον τρόπο που προοριζόταν από τον δημιουργό της έρευνας σχετίζεται αναπόσπαστα με την εγκυρότητα της έρευνας και τα συμπεράσματα που μπορούν να εξαχθούν από αυτή.

▪ **Βήμα 7: Διεξαγωγή Δοκιμών**

Εάν ο στόχος του ερωτηματολογίου αφορά σε εκτεταμένο δείγμα τότε κρίνεται απαραίτητη και η διεξαγωγή πιλοτικών δοκιμών του ερωτηματολογίου. είναι η πιλοτική δοκιμή του ερωτηματολογίου και η συνέχιση της συλλογής στοιχείων εγκυρότητας. Δύο από τις πιο κοινές προσεγγίσεις βασίζονται στην εσωτερική δομή και τις σχέσεις με άλλες μεταβλητές (Bichi et al., 2015) . Κατά τη διάρκεια των πιλοτικών δοκιμών, τα μέλη του πληθυσμού-στόχου ολοκληρώνουν την έρευνα με τον προγραμματισμένο τρόπο παράδοσης (π.χ. μορφή web ή έντυπη). Τα δεδομένα που προέκυψαν από την πιλοτική δοκιμή επανεξετάζονται στη συνέχεια για να αξιολογηθεί το εύρος και η διακύμανση των στοιχείων, η αξιοπιστία της βαθμολογίας και να επανεξεταστούν οι συσχετίσεις στοιχείων και σύνθετης βαθμολογίας. Κατά τη διάρκεια αυτού του βήματος, οι σχεδιαστές ερευνών θα πρέπει επίσης να επανεξετάσουν περιγραφικά στατιστικά στοιχεία (π.χ. μέσα και τυπικές αποκλίσεις) και ιστογράμματα, τα οποία δείχνουν την κατανομή των απαντήσεων ανά στοιχείο. Αυτή η ανάλυση μπορεί να βοηθήσει στον εντοπισμό στοιχείων που μπορεί να μην λειτουργούν με τον τρόπο που σκόπευε ο σχεδιαστής.

Για να εξακριβωθεί η εσωτερική δομή του ερωτηματολογίου και να αξιολογηθεί ο βαθμός στον οποίο τα στοιχεία μιας συγκεκριμένης κλίμακας μετρούν ένα κατασκευάσμα οι σχεδιαστές της έρευνας θα πρέπει να εξετάσουν το ενδεχόμενο χρήσης προηγμένων στατιστικών τεχνικών όπως η παραγοντική ανάλυση. Η παραγοντική ανάλυση είναι μια στατιστική διαδικασία που έχει σχεδιαστεί για να αξιολογήσει «τον αριθμό των διακριτών δομών που απαι-

τούνται για να ληφθεί υπόψη το πρότυπο συσχετίσεων μεταξύ ενός συνόλου μέτρων» (Fabrigar, et al., 2012). Οι ερευνητές θα πρέπει να γνωρίζουν ότι οι τεχνικές παραγοντικής ανάλυσης συχνά δεν είναι κατανοητές και εφαρμόζονται ελάχιστα αλλά, η βιβλιογραφία είναι γεμάτη με πολλούς χρήσιμους οδηγούς [19].

Η διεξαγωγή ανάλυσης αξιοπιστίας είναι ένα άλλο κρίσιμο βήμα στη φάση της πιλοτικής δοκιμής. Ο πιο συνηθισμένος τρόπος αξιολόγησης της αξιοπιστίας της κλίμακας είναι ο υπολογισμός του συντελεστή Alpha-Cronbach. Το Alpha-Cronbach είναι ένα μέτρο της εσωτερικής συνέπειας των βαθμολογιών των στοιχείων. Είναι συνάρτηση των συσχετίσεων μεταξύ των στοιχείων και του συνολικού αριθμού στοιχείων σε μια συγκεκριμένη κλίμακα. Είναι σημαντικό να σημειωθεί ότι το άλφα του Cronbach δεν είναι ένα καλό μέτρο της μονοδιάστατης κλίμακας (μέτρηση μιας μεμονωμένης έννοιας) όπως συχνά υποτίθεται (Schmitt, 1996).

Η προαναφερθείσα μεθοδολογία σχεδιασμού της έρευνας που παρουσιάζεται εδώ δεν είναι ο μόνος τρόπος σχεδιασμού και ανάπτυξης ενός ερωτηματολογίου υψηλής ποιότητας. Ωστόσο, αποτελεί μια συστηματική έρευνα, βασισμένη σε στοιχεία προσέγγισης για το σχεδιασμό ερωτηματολογίων. Κάτι τέτοιο όχι μόνο μπορεί να βελτιώσει τα ερωτηματολόγια που χρησιμοποιούνται στην εκπαίδευση, αλλά έχει επίσης τη δυνατότητα να επηρεάσει θετικά τη συνολική ποιότητα της εκπαιδευτικής έρευνας, όπου ένα η μερίδα του λέοντος της οποίας χρησιμοποιεί ερωτηματολόγια.

3.2 Ερωτηματολόγιο μαθητών

Τα συγκεκριμένα δύο ερωτηματολόγια, που αναπτύχθηκαν στο πλαίσιο της παρούσας διπλωματικής δομήθηκαν με τέτοιο τρόπο ώστε να ικανοποιούνται τα κριτήρια. Λήφθηκε υπόψη η δυσκολία των ερωτημάτων να είναι διαβαθμισμένη, να εμπεριέχονται ερωτήσεις οι οποίες να αφορούν πολύπλευρα την ύλη που έχουν διδαχτεί οι μαθητές της Α' γυμνασίου κατά τα σχολικά έτη φοίτησης του δημοτικού ενώ στους πρωτοετείς φοιτητές το ερωτηματολόγιο περιείχε ερωτήσεις αναφορικά με τις γνώσεις που έχουν λάβει όλα τα μαθητικά τους χρόνια ανεξάρτητα την ομάδα προσανατολισμού τους στο Λύκειο.

Συνολικά διανεμήθηκαν n=142 έντυπα ερωτηματολόγια των 25 ερωτήσεων που απευθύνονται σε αποφοίτους Στ' δημοτικού (και πιο συγκεκριμένα μαθητές Α' Γυμνασίου) εκ των οποίων διατηρήθηκαν 123 εξαιτίας των Missing Values (ποσοστό $\geq 50\%$). Τα ερωτηματολόγια διανεμήθηκαν στα εξής σχολεία: 7ο Γυμνάσιο Ιωαννίνων, 2ο Γυμνάσιο Ανατολής, Γυμνάσιο Λογγάδων. Σχετικά με το περιεχόμενο του ερωτηματολογίου λήφθηκαν υπόψη όλες

οι ενότητες των μαθηματικών της Στ' Δημοτικού ώστε να μην υπάρχει μεροληψία υπέρ ή έναντι τμήματος διδακτέας ύλης.

Πιο συγκεκριμένα οι ενότητες αφορούν:

- A.** Θεματική Ενότητα 1. Αριθμοί και πράξεις
- B.** Θεματική Ενότητα 2. Εξισώσεις
- Γ.** Θεματική Ενότητα 3. Συλλογή και επεξεργασία δεδομένων
- Δ.** Θεματική Ενότητα 4. Μετρήσεις - Μοτίβα
- Ε.** Θεματική Ενότητα 5. Γεωμετρία

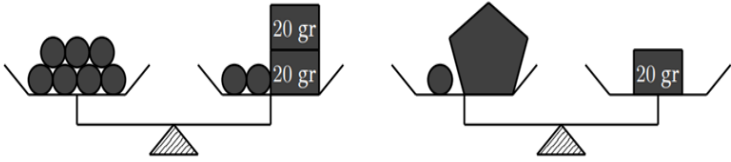
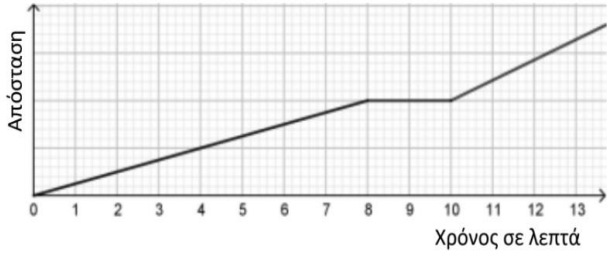
Επιπλέον, κάποια από τα ερωτήματα λήφθηκαν από Θέματα γραπτής δοκιμασίας εισαγωγής μαθητών στα Π.Π. Γυμνάσια καθώς και την Κυπριακή Μαθηματική Ολυμπιάδα για μαθητές Στ' Δημοτικού.

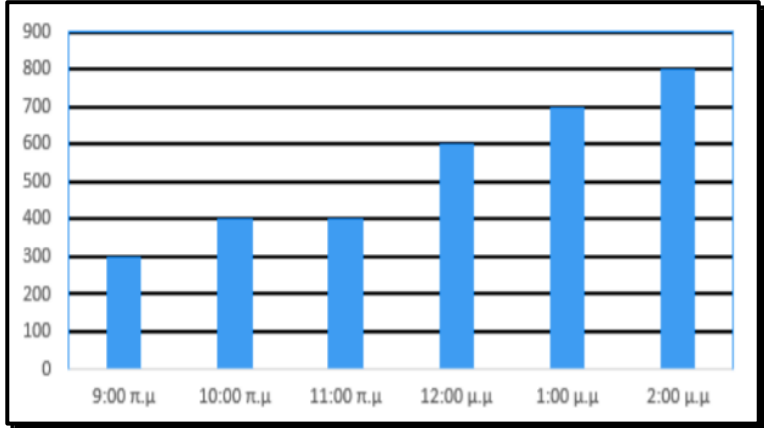
Τα Μαθηματικά...σε ερωτήσεις

Φύλο: Κορίτσι Αγόρι

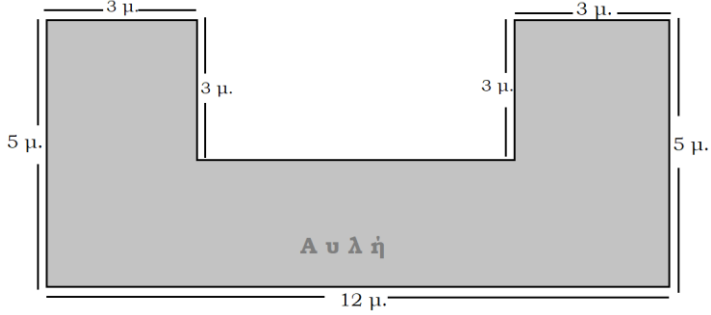
Κύκλωσε την απάντηση που θεωρείς ότι είναι η σωστή. Μπορείς να κυκλώσεις μόνο μία.

1.	<p>Στην εξίσωση $x \cdot (3^2 : 3) = 30$, βρείτε το x:</p> <p>A. $x = 15$ B. $x = 8$ Γ. $x = 10$ Δ. $x = 1$ E. Δε γνωρίζω</p>
2.	<p>Ποιος από τους παρακάτω αριθμούς είναι πιο κοντά στο 9;</p> <p>A. 9,9 B. 9,99 Γ. 10,009 Δ. 10,01 E. Δε γνωρίζω</p>
3.	<p>Ποιο από τα παρακάτω κλάσματα βρίσκεται ανάμεσα στο $\frac{1}{5}$ και στο $\frac{2}{3}$.</p> <p>A. $\frac{4}{3}$ B. $\frac{9}{15}$ Γ. $\frac{5}{6}$ Δ. $\frac{1}{10}$ E. Δε γνωρίζω</p>
4.	<p>Με $1\frac{1}{2}$ λίτρα γάλα πόσα ποτήρια του $\frac{1}{4}$ λίτρου γεμίζουμε;</p> <p>A. 4 ποτήρια B. 6 ποτήρια Γ. $4\frac{1}{4}$ ποτήρια Δ. 5 ποτήρια E. Δε γνωρίζω</p>
5.	<p>Ο Ερμής έχει δυο κομπιουτεράκια, το Α και το Β. Το Α κάνει 17 πολλαπλασιασμούς ανά δευτερόλεπτο ενώ το Β κάνει 1017 πολλαπλασιασμούς ανά λεπτό. Πόσους περισσότερους πολλαπλασιασμούς κάνει το Α από το Β σε δυο λεπτά;</p> <p>A. 10 B. 9 Γ. 6 Δ. 3 E. Δε γνωρίζω</p>
6.	<p>Ποιον αριθμό θα βρω, αν το 5,94, το πολλαπλασιάσω με το 10, έπειτα, αφαιρέσω τον αριθμό 9 και, τέλος, διαιρέσω τον αριθμό που θα βρω με το 0,1;</p> <p>A. 50,4 B. 40,4 Γ. 404 Δ. 504 E. Δε γνωρίζω</p>
7.	<p>Ένας αριθμός Α αναλύθηκε σε γινόμενο με τον παρακάτω τρόπο: $A=2 \times 3 \times 3 \times 17$. Το πλήθος των διαιρετών του είναι:</p> <p>A. 3 B. 4 Γ. 5 Δ. 16 E. Δε γνωρίζω</p>
8.	<p>Το γινόμενο ενός φυσικού αριθμού Α επί το διπλάσιο του ισούται με 72. Το μισό του Α ισούται με:</p> <p>A. 3 B. 6 Γ. 18 Δ. 36 E. Δε γνωρίζω</p>

9.	<p>Ο διευθυντής ενός σχολείου προμηθεύτηκε 47 εισιτήρια λεωφορείου για τη μετάβαση των μαθητών στον αρχαιολογικό χώρο της Κνωσού και 47 εισιτήρια για την επιστροφή τους. Επειδή κάποιοι μαθητές απουσίαζαν δε χρησιμοποιήθηκαν 6 εισιτήρια. Με ποια εξίσωση μπορούμε να βρούμε τον αριθμό των μαθητών που πήγαν στην Κνωσό;</p> <p>A. $x+6=94$ B. $94+6=x \cdot 2$ Γ. $2 \cdot x+6=94$ Δ. $94-x=6$ Ε. Δε γνωρίζω</p>
10.	<p>Αν το άθροισμα των αριθμών των δύο τελευταίων σελίδων ενός βιβλίου είναι 155 τότε πόσες σελίδες έχει το βιβλίο;</p> <p>A. 77 B. 78 Γ. 155 Δ. Περισσότερες από 155 Ε. Δε γνωρίζω</p>
11.	<p>Στις δύο ζυγαριές υπάρχουν σφαιρικοί βόλοι, ένα πεντάγωνο και σταθμά των 20 γρ. Αν οι ζυγαριές ισορροπούν, τότε στο πεντάγωνο αντιστοιχούν:</p> <div style="text-align: center;">  </div> <p>A. 0,010 κιλά B. 0,011κιλά Γ. 0,012 κιλά Δ. 0,013 κιλά Ε. Δε γνωρίζω</p>
12.	<p>Στην ισότητα $\text{☺} + \text{☺} + 5 + \text{☺} = \text{☺} + 12 + \text{☺}$ ο αριθμός ☺ είναι ίσος με:</p> <p>A. 5 B. 12 Γ. 17 Δ. 7 Ε. Δε γνωρίζω</p>
13.	<p>Η γραφική παράσταση παρουσιάζει τη διαδρομή που κάνει ο Μάριος με το σχολικό λεωφορείο. Αν μπήκε στο λεωφορείο στις 7:40, τι ώρα ήταν, όταν αυτό σταμάτησε για να πάρει ακόμα δύο συμμαθητές του;</p> <div style="text-align: center;">  </div> <p>A. 7:44 B. 7:48 Γ. 7:52 Δ. 7:54 Ε. Δε γνωρίζω</p>
14.	<p>Σε μια εταιρεία 3 υπάλληλοι χρειάζονται 18 ημέρες για να ολοκληρώσουν μια εργασία. Σε πόσες ημέρες θα ολοκληρώσουν την εργασία αν εργαστούν 9 υπάλληλοι;</p> <p>A. σε 54 ημέρες B. σε 6 ημέρες Γ. σε 27 ημέρες Δ. σε 3 ημέρες Ε. Δε γνωρίζω</p>
15.	<p>Ένα πλυντήριο κοστίζει 800€. Πάνω του υπάρχει καρτελάκι που γράφει τιμή πώλησης με κέρδος 25% στο κόστος. Στις εκπτώσεις ο καταστηματάρχης βάζει έκπτωση 25% σε όλα τα προϊόντα. Το πλυντήριο τότε πωλείται:</p>

	A. 750 B. 650 Γ. 400 Δ. 300 Ε. Δε γνωρίζω															
16.	<p>Το 25% των μαθητών του σχολείου παίζει ποδόσφαιρο. Αν οι μαθητές του σχολείου, που δεν παίζουν ποδόσφαιρο, είναι 180 τότε το σχολείο πόσους μαθητές έχει;</p> <p>A. 240 B. 225 Γ. 135 Δ. 250 Ε. Δε γνωρίζω</p>															
17.	<p>Στον παρακάτω πίνακα φαίνονται οι επιδόσεις 4 παικτών μιας ομάδας μπάσκετ, κατά τη διάρκεια ενός αγώνα, σε τρίποντα.</p> <table border="1" data-bbox="518 560 1117 896"> <thead> <tr> <th>Παίκτης</th> <th>Προσπάθειες για τρίποντο</th> <th>Επιτυχίες</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>10</td> <td>5</td> </tr> <tr> <td>B</td> <td>20</td> <td>5</td> </tr> <tr> <td>Γ</td> <td>16</td> <td>8</td> </tr> <tr> <td>Δ</td> <td>10</td> <td>8</td> </tr> </tbody> </table> <p>Εάν ήσουν προπονητής/τρια της ομάδας αυτής ποιον θα επέλεγες για να εκτελέσει το τελευταίο και κρίσιμο τρίποντο του αγώνα;</p> <p>A. Τον A B. Τον B Γ. Τον Γ Δ. Τον Δ Ε. Δε γνωρίζω</p>	Παίκτης	Προσπάθειες για τρίποντο	Επιτυχίες	A	10	5	B	20	5	Γ	16	8	Δ	10	8
Παίκτης	Προσπάθειες για τρίποντο	Επιτυχίες														
A	10	5														
B	20	5														
Γ	16	8														
Δ	10	8														
18.	<p>Το παρακάτω διάγραμμα αναπαριστά τον ημερήσιο αριθμό των επισκεπτών σε ένα μουσείο ανά ώρα από τις 9:00 π.μ. έως τις 2:00 μ.μ. Το ποσοστό αύξησης στα εκατό του αριθμού των επισκεπτών από τις 11:00 π.μ. έως τις 12 μ.μ. είναι</p>  <table border="1" data-bbox="430 1198 1197 1624"> <thead> <tr> <th>Ωρα</th> <th>Αριθμός επισκεπτών</th> </tr> </thead> <tbody> <tr> <td>9:00 π.μ.</td> <td>300</td> </tr> <tr> <td>10:00 π.μ.</td> <td>400</td> </tr> <tr> <td>11:00 π.μ.</td> <td>400</td> </tr> <tr> <td>12:00 μ.μ.</td> <td>600</td> </tr> <tr> <td>1:00 μ.μ.</td> <td>700</td> </tr> <tr> <td>2:00 μ.μ.</td> <td>800</td> </tr> </tbody> </table> <p>A. 33% B. 50% Γ. 200% Δ. 20% Ε. Δε γνωρίζω</p>	Ωρα	Αριθμός επισκεπτών	9:00 π.μ.	300	10:00 π.μ.	400	11:00 π.μ.	400	12:00 μ.μ.	600	1:00 μ.μ.	700	2:00 μ.μ.	800	
Ωρα	Αριθμός επισκεπτών															
9:00 π.μ.	300															
10:00 π.μ.	400															
11:00 π.μ.	400															
12:00 μ.μ.	600															
1:00 μ.μ.	700															
2:00 μ.μ.	800															
19.	<p>Το ραβδόγραμμα δείχνει τον αριθμό των απαντήσεων των παιδιών με θέμα την αγαπημένη ενασχόλησή τους στον ελεύθερο χρόνο. Τι ποσοστό των παιδιών απάντησαν ότι η αγαπημένη τους ενασχόληση είναι η μουσική;</p>															

	<p style="text-align: center;">Ενασχόληση στον ελεύθερο χρόνο</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Ενασχόληση</th> <th>Ποσοστό (%)</th> </tr> </thead> <tbody> <tr> <td>ηλεκτρονικά παιχνίδια</td> <td>30</td> </tr> <tr> <td>παιχνίδι με φίλους</td> <td>20</td> </tr> <tr> <td>μουσική</td> <td>20</td> </tr> <tr> <td>λογοτεχνία</td> <td>10</td> </tr> </tbody> </table> <p>A. 25% B. 20% Γ. 30% Δ. 28% Ε. Δε γνωρίζω</p>	Ενασχόληση	Ποσοστό (%)	ηλεκτρονικά παιχνίδια	30	παιχνίδι με φίλους	20	μουσική	20	λογοτεχνία	10																	
Ενασχόληση	Ποσοστό (%)																											
ηλεκτρονικά παιχνίδια	30																											
παιχνίδι με φίλους	20																											
μουσική	20																											
λογοτεχνία	10																											
<p>20.</p>	<p>Στο παρακάτω μοτίβο που επαναλαμβάνεται, παρατηρούμε δύο κομμάτια πίτσα στην 4^η σειρά. Σε ποια σειρά θα εμφανιστούν ξανά τα δύο κομμάτια πίτσα;</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>1^η</th> <th>2^η</th> <th>3^η</th> <th>4^η</th> <th>5^η</th> <th>6^η</th> <th>7^η</th> <th>8^η</th> <th>9^η</th> </tr> </thead> <tbody> <tr> <td>▽</td> <td>○</td> <td>○</td> <td>▽</td> <td>○</td> <td>○</td> <td>▽</td> <td>○</td> <td>○</td> </tr> <tr> <td>○</td> <td>○</td> <td>○</td> <td>▽</td> <td>○</td> <td>○</td> <td>○</td> <td>▽</td> <td>○</td> </tr> </tbody> </table> <p>A. 10^η B. 16^η Γ. 14^η Δ. 17^η Ε. Δε γνωρίζω</p>	1 ^η	2 ^η	3 ^η	4 ^η	5 ^η	6 ^η	7 ^η	8 ^η	9 ^η	▽	○	○	▽	○	○	▽	○	○	○	○	○	▽	○	○	○	▽	○
1 ^η	2 ^η	3 ^η	4 ^η	5 ^η	6 ^η	7 ^η	8 ^η	9 ^η																				
▽	○	○	▽	○	○	▽	○	○																				
○	○	○	▽	○	○	○	▽	○																				
<p>21.</p>	<p>Δίνονται τα παρακάτω μοτίβα:</p> <p>1^ο: 10, 14, 18, 22, 26,....</p> <p>2^ο: 4, 13, 22, 31, 40,....</p> <p>Μετά τον αριθμό 22 που είναι κοινός στα δύο μοτίβα, ο επόμενος κοινός αριθμός θα είναι:</p> <p>A. 58 B. 62 Γ. 61 Δ. 64 Ε. Δε γνωρίζω</p>																											
<p>22.</p>	<p>Η περίμετρος τετραγώνου με εμβαδό 36 τετραγωνικά μέτρα είναι:</p> <p>A. 6 μέτρα B. 24 μέτρα Γ. 36 μέτρα Δ. 144 μέτρα Ε. Δε γνωρίζω</p>																											
<p>23.</p>	<p>Στην αυλή της παρακάτω εικόνας, θα τοποθετηθεί ένας τάπητας. Αν η τιμή του τάπητα είναι 10€/τ.μ., το συνολικό κόστος θα είναι:</p>																											

	 <p style="text-align: center;">Αυλή</p> <p style="text-align: center;">12 μ.</p> <p>A. 240 ευρώ B. 420 ευρώ Γ. 600 ευρώ Δ. 960 ευρώ E. Δε γνωρίζω</p>
24.	<p>Σε ένα τρίγωνο ΑΒΓ η γωνία Α είναι 40 μοίρες. Τότε το τρίγωνο είναι:</p> <p>A. Αμβλυγώνιο B. Οξυγώνιο Γ. Ορθογώνιο Δ. Δεν αρκεί η δοθείσα πληροφορία E. Δε γνωρίζω</p>
25.	<p>Ένας ξύλινος κύβος με πλευρά 10 εκ. θα κοπεί σε κύβους πλευράς 3 εκ. ο καθένας. Ο μέγιστος αριθμός κύβων, πλευράς 3 εκ., που μπορούν να δημιουργηθούν είναι:</p> <p>A. 37 B. 36 Γ. 27 Δ. 26 E. Δε γνωρίζω</p>

3.3 Ερωτηματολόγιο Φοιτητών

Συνολικά διανεμήθηκαν $n=99$ έντυπα ερωτηματολόγια των 32 ερωτήσεων που απευθύνονται σε αποφοίτους Λυκείου (και πιο συγκεκριμένα φοιτητές που παρακολουθούν το μάθημα της Διδακτικής των Μαθηματικών - 1ου Εξαμήνου του ΠΤΔΕ) εκ των οποίων διατηρήθηκαν 98 εξαιτίας των Missing Values (ποσοστό $\geq 50\%$) κριτήριο το οποίο τέθηκε κατά την ανάλυση των δεδομένων. Στα συγκεκριμένα ερωτηματολόγια υπάρχουν 25 κοινά ερωτήματα με το Ερωτηματολόγιο των μαθητών.

Τα Μαθηματικά...σε ερωτήσεις

Αγαπητοί φοιτητές/ριες του Π.Τ.Δ.Ε.,

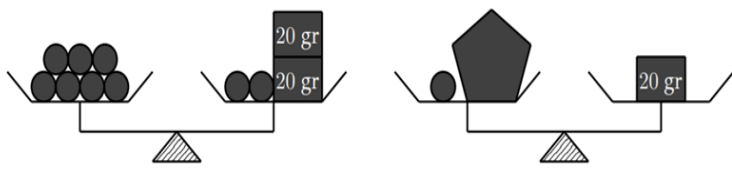
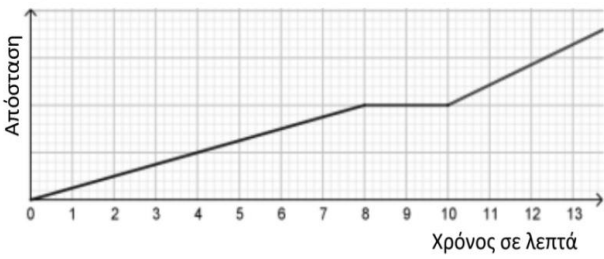
Το παρόν ερωτηματολόγιο είναι ανώνυμο, εμπιστευτικό και τα αποτελέσματα θα χρησιμοποιηθούν αυστηρά και μόνο στο πλαίσιο της στατιστικής ανάλυσης της διπλωματικής εργασίας μεταπτυχιακού φοιτητή του τμήματος. Θα χρειαστείτε το πολύ μία ώρα για την συμπλήρωσή του. Η συμβολή σας είναι ιδιαίτερα σημαντική. Ευχαριστούμε για τον χρόνο σας.

Φύλο:	Θήλυ <input type="checkbox"/>			Άρρεν <input type="checkbox"/>	
Έτος Σπουδών:	1 ^ο <input type="checkbox"/>	2 ^ο <input type="checkbox"/>	3 ^ο <input type="checkbox"/>	4 ^ο <input type="checkbox"/>	μεγαλύτερο <input type="checkbox"/>

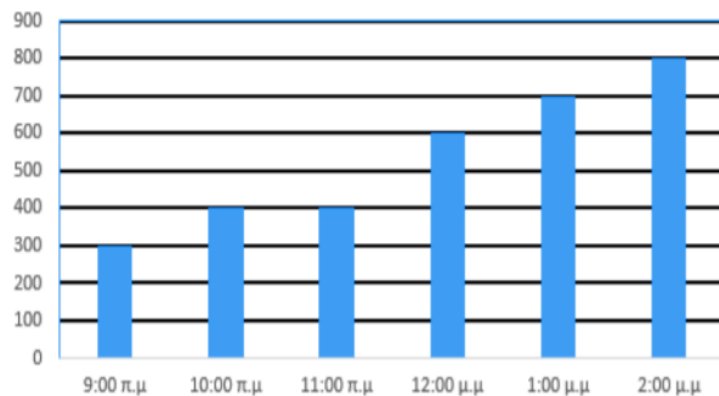
Κύκλωσε την απάντηση που θεωρείς ότι είναι η σωστή. Μπορείς να κυκλώσεις μόνο μία.

1.	Στην εξίσωση $x \cdot (3^2 : 3) = 30$, βρείτε το x : A. $x = 15$ B. $x = 8$ Γ. $x = 10$ Δ. $x = 1$ Ε. Δε γνωρίζω
2.	Ποιος από τους παρακάτω αριθμούς είναι πιο κοντά στο 9; A. 9,9 B. 9,99 Γ. 10,009 Δ. 10,01 Ε. Δε γνωρίζω
3.	Ποιο από τα παρακάτω κλάσματα βρίσκεται ανάμεσα στο $\frac{1}{5}$ και στο $\frac{2}{3}$.

	A. $\frac{4}{3}$ B. $\frac{9}{15}$ Γ. $\frac{5}{6}$ Δ. $\frac{1}{10}$ Ε. Δε γνωρίζω
4.	Με $1\frac{1}{2}$ λίτρα γάλα πόσα ποτήρια του $\frac{1}{4}$ λίτρου γεμίζουμε; A. 4 ποτήρια B. 6 ποτήρια Γ. $4\frac{1}{4}$ ποτήρια Δ. 5 ποτήρια Ε. Δε γνωρίζω
5.	Ο Ερμής έχει δυο κομπιουτεράκια, το Α και το Β. Το Α κάνει 17 πολλαπλασιασμούς ανά δευτερόλεπτο ενώ το Β κάνει 1017 πολλαπλασιασμούς ανά λεπτό. Πόσους περισσότερους πολλαπλασιασμούς κάνει το Α από το Β σε δυο λεπτά; A. 10 B. 9 Γ. 6 Δ. 3 Ε. Δε γνωρίζω
6.	Ποιον αριθμό θα βρω, αν το 5,94, το πολλαπλασιάσω με το 10, έπειτα, αφαιρέσω τον αριθμό 9 και, τέλος, διαιρέσω τον αριθμό που θα βρω με το 0,1; A. 50,4 B. 40,4 Γ. 404 Δ. 504 Ε. Δε γνωρίζω
7.	Ένας αριθμός Α αναλύθηκε σε γινόμενο με τον παρακάτω τρόπο: $A=2 \times 3 \times 3 \times 17$. Το πλήθος των διαιρετών του είναι: A.3 B.4 Γ.5 Δ. 16 Ε. Δε γνωρίζω
8.	Το γινόμενο ενός φυσικού αριθμού Α επί το διπλάσιο του ισούται με 72. Το μισό του Α ισούται με: A.3 B.6 Γ.18 Δ.36 Ε. Δε γνωρίζω
9.	Ο διευθυντής ενός σχολείου της Κρήτης προμηθεύτηκε 47 εισιτήρια λεωφορείου για τη μετάβαση των μαθητών της Στ΄ τάξης στον αρχαιολογικό χώρο της Κνωσού και 47 εισιτήρια για την επιστροφή τους. Επειδή κάποιοι μαθητές απουσίαζαν δε χρησιμοποιήθηκαν 6 εισιτήρια. Με ποια εξίσωση μπορούμε να βρούμε τον αριθμό των

	<p>μαθητών που πήγαν στην Κνωσό;</p> <p>A. $x+6=94$ B. $94+6=x \cdot 2$ Γ. $2 \cdot x+6=94$ Δ. $94-x=6$ Ε. Δε γνωρίζω</p>
10.	<p>Αν το άθροισμα των αριθμών των δύο τελευταίων σελίδων ενός βιβλίου είναι 155 τότε πόσες σελίδες έχει το βιβλίο έχει;</p> <p>A. 77 B. 78 Γ. 155 Δ. Περισσότερες από 155 Ε. Δε γνωρίζω</p>
11.	<p>Στις δύο ζυγαριές υπάρχουν σφαιρικοί βόλοι, ένα πεντάγωνο και σταθμά των 20 γρ. Αν οι ζυγαριές ισορροπούν, τότε στο πεντάγωνο αντιστοιχούν:</p>  <p>A. 0,010 κιλά B. 0,011κιλά. Γ. 0,012 κιλά. Δ. 0,013 κιλά. Ε. Δε γνωρίζω</p>
12.	<p>Στην ισότητα $\text{☺} + \text{☺} + 5 + \text{☺} = \text{☺} + 12 + \text{☺}$ ο αριθμός ☺ είναι ίσος με:</p> <p>A. 5 B. 12 Γ. 17 Δ. 7 Ε. Δε γνωρίζω</p>
13.	<p>Η γραφική παράσταση παρουσιάζει τη διαδρομή που κάνει ο Μάριος με το σχολικό λεωφορείο. Αν μπήκε στο λεωφορείο στις 7:40, τι ώρα ήταν, όταν αυτό σταμάτησε για να πάρει ακόμα δύο συμμαθητές του;</p>  <p>A. 7:44 B. 7:48 Γ. 7:52 Δ. 7:54 Ε. Δε γνωρίζω</p>

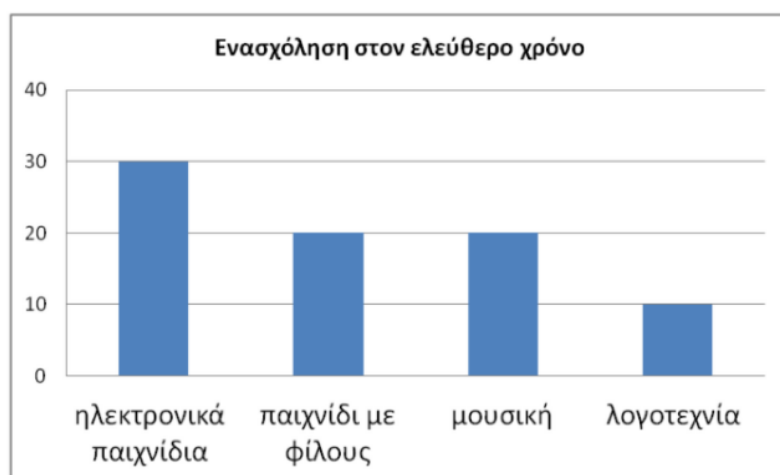
14.	<p>Σε μια εταιρεία 3 υπάλληλοι χρειάζονται 18 ημέρες για να ολοκληρώσουν μια εργασία. Σε πόσες ημέρες θα ολοκληρώσουν την εργασία αν εργαστούν 9 υπάλληλοι;</p> <p>A. σε 54 ημέρες B. σε 6 ημέρες Γ. σε 27 ημέρες Δ. σε 3 ημέρες Ε. Δε γνωρίζω</p>															
15.	<p>Ένα πλυντήριο κοστίζει 800€. Πάνω του υπάρχει καρτελάκι που γράφει τιμή πώλησης με κέρδος 25% στο κόστος. Στις εκπτώσεις ο καταστηματάρχης βάζει έκπτωση 25% σε όλα τα προϊόντα. Το πλυντήριο τότε πωλείται:</p> <p>A. 750 B. 650 Γ. 400 Δ. 300 Ε. Δε γνωρίζω</p>															
16.	<p>Το 25% των μαθητών του σχολείου παίζει βόλεϊ. Αν οι μαθητές του σχολείου που δεν παίζουν βόλεϊ είναι 180 τότε το σχολείο πόσους μαθητές έχει;</p> <p>A. 240 B. 225 Γ. 135 Δ. 250 Ε. Δε γνωρίζω</p>															
17.	<p>Στον παρακάτω πίνακα φαίνονται οι επιδόσεις 5 παικτών μιας ομάδας μπάσκετ, κατά τη διάρκεια ενός αγώνα, σε τρίποντα.</p> <table border="1" data-bbox="499 1133 1134 1469"> <thead> <tr> <th>Παίκτης</th> <th>Προσπάθειες για τρίποντο</th> <th>Επιτυχίες</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>10</td> <td>5</td> </tr> <tr> <td>B</td> <td>20</td> <td>5</td> </tr> <tr> <td>Γ</td> <td>16</td> <td>8</td> </tr> <tr> <td>Δ</td> <td>10</td> <td>8</td> </tr> </tbody> </table> <p>Εάν ήσουν προπονητής/τρια της ομάδας αυτής ποιον θα επέλεγες για να εκτελέσει το τελευταίο και κρίσιμο τρίποντο του αγώνα;</p> <p>A. A B. B Γ. Γ Δ. Δ Ε. Δε γνωρίζω</p>	Παίκτης	Προσπάθειες για τρίποντο	Επιτυχίες	A	10	5	B	20	5	Γ	16	8	Δ	10	8
Παίκτης	Προσπάθειες για τρίποντο	Επιτυχίες														
A	10	5														
B	20	5														
Γ	16	8														
Δ	10	8														
18.	<p>Το παρακάτω διάγραμμα αναπαριστά τον ημερήσιο αριθμό των επισκεπτών σε ένα μουσείο ανά ώρα από τις 9:00 π.μ. έως τις 2:00 μ.μ. Το ποσοστό αύξησης στα εκατό του αριθμού των επισκεπτών από τις 11:00 π.μ. έως τις 12 μ.μ. είναι</p>															



Α.33% Β.50% Γ.200% Δ.20% Ε. Δε γνωρίζω

19.

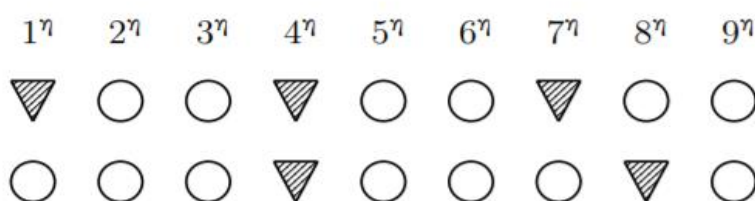
Το ραβδόγραμμα δείχνει τον αριθμό των απαντήσεων των παιδιών με θέμα την αγαπημένη ενασχόλησή τους στον ελεύθερο χρόνο. Τι ποσοστό των παιδιών απάντησαν ότι η αγαπημένη τους ενασχόληση είναι η μουσική;



Α.25% Β.20% Γ.30% Δ.28% Ε. Δε γνωρίζω

20.

Στο παρακάτω μοτίβο που επαναλαμβάνεται, παρατηρούμε δύο κομμάτια πίτσα στην 4η σειρά. Σε ποια σειρά θα εμφανιστούν ξανά τα δύο κομμάτια πίτσα;



Α.10η Β.16η Γ.14η Δ.17η Ε. Δε γνωρίζω

21.	<p>Δίνονται τα παρακάτω μοτίβα:</p> <p>1^ο: 10, 14, 18, 22, 26,....</p> <p>2^ο: 4, 13, 22, 31, 40,....</p> <p>Μετά τον αριθμό 22 που είναι κοινός και στα δύο μοτίβα, ο επόμενος κοινός αριθμός θα είναι:</p> <p>A.58 B.62 Γ.61 Δ.64 E. Δε γνωρίζω</p>
22.	<p>Η περίμετρος τετραγώνου με εμβαδό 36 τετραγωνικά μέτρα είναι:</p> <p>A. 6 μέτρα B. 24 μέτρα Γ. 36 μέτρα Δ. 144 μέτρα E. Δε γνωρίζω</p>
23.	<p>Στην αυλή της παρακάτω εικόνας, θα τοποθετηθεί ένας τάπητας. Αν η τιμή του τάπητα είναι 10€/τ.μ., το συνολικό κόστος θα είναι:</p> <div style="text-align: center;"> </div> <p>A. 240 ευρώ B. 420 ευρώ Γ. 600 ευρώ Δ. 960 ευρώ E. Δε γνωρίζω</p>
24.	<p>Σε ένα τρίγωνο ΑΒΓ η γωνία Α είναι 40 μοίρες. Τότε το τρίγωνο είναι:</p> <p>A. Αμβλυγώνιο B. οξυγώνιο Γ. ορθογώνιο Δ. Δεν αρκεί η δοθείσα πληροφορία E. Δε γνωρίζω</p>
25.	<p>Ένας ξύλινος κύβος με πλευρά 10 εκ. θα κοπεί σε κύβους πλευράς 3 εκ. ο καθένας. Ο μέγιστος αριθμός κύβων, πλευράς 3 εκ., που μπορούν να δημιουργηθούν είναι:</p>

	A.37 B. 36 Γ. 27 Δ.26 E. Δε γνωρίζω
26.	<p>Ο Ορφέας αποφάσισε να ασχοληθεί με το τρέξιμο. Αρχικά, έτρεχε τα 15 χιλιόμετρα σε 1 ώρα και 30 λεπτά. Έπειτα από αρκετές προπονήσεις τρέχει τα 24 χιλιόμετρα σε 2 ώρες. Τώρα πια σε σχέση με την αρχική του επίδοση, το ένα χιλιόμετρο το διανύει σε:</p> <p>A. 1 λεπτό λιγότερο B.2 λεπτά λιγότερο Γ. 3 λεπτά λιγότερο Δ. 4 λεπτά λιγότερο E. Δε γνωρίζω</p>
27.	<p>Ένας τριψήφιος αριθμός είναι πολλαπλάσιο του 5 και του 8. Αν το άθροισμα των ψηφίων του είναι 3, τότε το γινόμενο των ψηφίων του είναι:</p> <p>A.0 B.40 Γ.120 Δ.16 E. Δε γνωρίζω</p>
28.	<p>Σε ένα σπίτι συγκεντρώθηκαν κάποια άτομα και αντάλλαξαν όλοι με όλους χειραψίες. Αν οι χειραψίες ήταν 28 πόσα άτομα συγκεντρώθηκαν;</p> <p>A.8 B.14 Γ.27 Δ.28 E. Δε γνωρίζω</p>
29.	<p>Ο Παναγιώτης είναι 7 χρόνια μεγαλύτερος από την Αφροδίτη. Η Αφροδίτη είναι 7 χρόνια μικρότερη από την Ευαγγελία. Αν ο Παναγιώτης είναι 33 ετών, πόσων ετών είναι η Ευαγγελία;</p> <p>A. 19 B. 33 Γ. 40 Δ. 47 E. Δε γνωρίζω</p>
30.	<p>Ποιος είναι ο αριθμός εκείνος, που, αν αφαιρέσω 10 από το μισό του, προκύπτει δι-αφορά 16;</p> <p>A 21 B. 52 Γ. 13 Δ. 18 E. Δε γνωρίζω</p>
31.	<p>Σε ένα επιτραπέζιο παιχνίδι ερωτήσεων ισχύει ο εξής κανόνας: σε κάθε σωστή απά-</p>

	<p>ντηση δίνονται δύο βαθμοί, ενώ σε κάθε λανθασμένη καταλογίζεται αφαίρεση ενός βαθμού. Σε έναν γύρο 10 ερωτήσεων του παιχνιδιού, ο Στέφανος απάντησε σε όλες τις ερωτήσεις και συνολικά συγκέντρωσε 14 βαθμούς. Συνεπώς ο Στέφανος πόσες λανθασμένες απαντήσεις έδωσε;</p> <p>A.4 B.5 Γ.8 Δ.2 Ε. Δε γνωρίζω</p>
<p>32.</p>	<p>Από τους 250 μαθητές ενός σχολείου το 40% μετακινείται με λεωφορείο, το 38% με αυτοκίνητο και οι υπόλοιποι με τα πόδια. Πόσα παιδιά έρχονται με τα πόδια στο σχολείο;</p> <p>A. 195 B. 25 Γ. 55 Δ. 95 Ε. Δε γνωρίζω</p>

ΚΕΦΑΛΑΙΟ 4

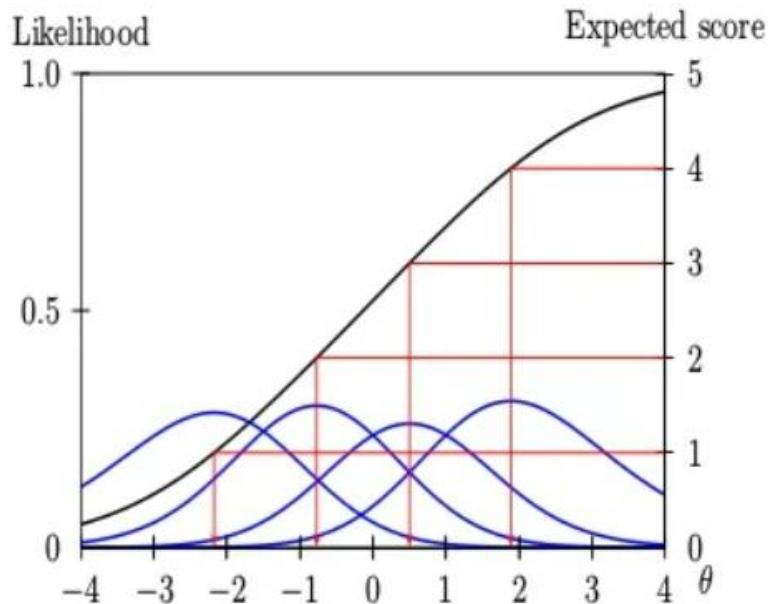
ΑΠΟΤΕΛΕΣΜΑΤΑ - ΣΥΜΠΕΡΑΣΜΑΤΑ

-
- 4.1 Αποτελέσματα
 - 4.2 Σύγκριση Μοντέλων
 - 4.3 Γραφικές
 - 4.4 Μελλοντική Έρευνα
-

Προκειμένου να χρησιμοποιήσουμε τα μοντέλα της IRT, για κάθε μαθητή και φοιτητή, αντιστοιχίσαμε σε κάθε ένα από τα 25 και τα 32 ερωτήματα αντίστοιχα την τιμή 1, αν το ερώτημα αυτό απαντήθηκε σωστά, και την τιμή 0 αν απαντήθηκε λάθος. Τα αρχεία δεδομένων που δημιουργήσαμε ήταν σε μορφή csv (comma-separated value) και κάθε φοιτητής/μαθητής αναπαραστάθηκε σαν ένα δυαδικό διάνυσμα μεγέθους 25 και 32. Άρα, το κάθε αρχείο των δεδομένων μας περιελάμβανε συνολικά 14 διανύσματα τέτοιου είδους και για την ανάλυση των δεδομένων αυτών χρησιμοποιήσαμε την ευρέως διαδεδομένη στατιστική γλώσσα προγραμματισμού R και την βιβλιοθήκη LTM (Latent Trait Models). Στο πακέτο αυτό περιέχονται συναρτήσεις οι οποίες προσπαθούν να ταιριάζουν το κάθε μοντέλο με διχοτομικά βαθμολογημένα δεδομένα χρησιμοποιώντας την μέθοδο της εκτίμησης της Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimate MLE) και γίνεται ενδελεχής αναφορά στο Παράρτημα.

Για να δώσουμε μία σύντομη ερμηνεία αυτής της μεθόδου εκτίμησης (Παξινού et al., 2017), αναφέρουμε ότι γνωρίζοντας τις απαντήσεις του φοιτητή στα ερωτήματα ενός τεστ, μπορούμε να βρούμε την ικανότητά του υπολογίζοντας την πιο πιθανή τιμή της ικανότητας που απαιτείται να έχει ο φοιτητής για να απαντήσει το 50% των ερωτημάτων. Αυτή είναι η έννοια της μεθόδου της εκτίμησης της μέγιστης πιθανοφάνειας. Έχουν ένα επαναληπτικό χαρακτήρα (iterative methods). Πρόκειται για επαναληπτική τεχνική και συνεπώς ξεκινά από μία αρχική εκτίμηση της δυσκολίας των ερωτημάτων στοχεύοντας να υπολογίσει την πιο πιθανή ικανότητα για κάθε εξεταζόμενο. Στη συνέχεια, με δεδομένη αυτήν την πιθανή ικανότητα του καθενός, γίνεται εκ νέου εκτίμηση της δυσκολίας των ερωτημάτων. Η διαδικασία αυτή

επαναλαμβάνεται έως ότου η απόσταση στις εκτιμήσεις των τιμών της δυσκολίας των ερωτημάτων, μεταξύ δύο διαδοχικών επαναλήψεων, να είναι μικρή, πράγμα το οποίο υποδηλώνει ότι έχει προκύψει η επιθυμητή σύγκλιση της επαναληπτικής αυτής μεθόδου.



Εικόνα 5 Βρίσκοντας την ικανότητα με την Μέγιστη Πιθανοφάνειας

Για να γίνουν πλήρως κατανοητοί οι πίνακες που ακολουθούν οφείλουμε να επισημάνουμε την ερμηνεία των παραμέτρων που παρουσιάζεται σε κάθε μοντέλο.

- **Παράμετρος Δυσκολίας:** Πόσο δύσκολη είναι μία ερώτηση
- **Παράμετρος Διάκρισης:** Πόσο καλά μπορεί μια ερώτηση (item) να διακρίνει την υψηλή και την χαμηλή ικανότητα του εξεταζόμενου
- **Παράμετρος Πρόβλεψης:** Αφορά κυρίως τον εξεταζόμενο με χαμηλή ικανότητα και μας ενημερώνει ποια η πιθανότητα να απαντήσει σωστά

4.1 Αποτελέσματα

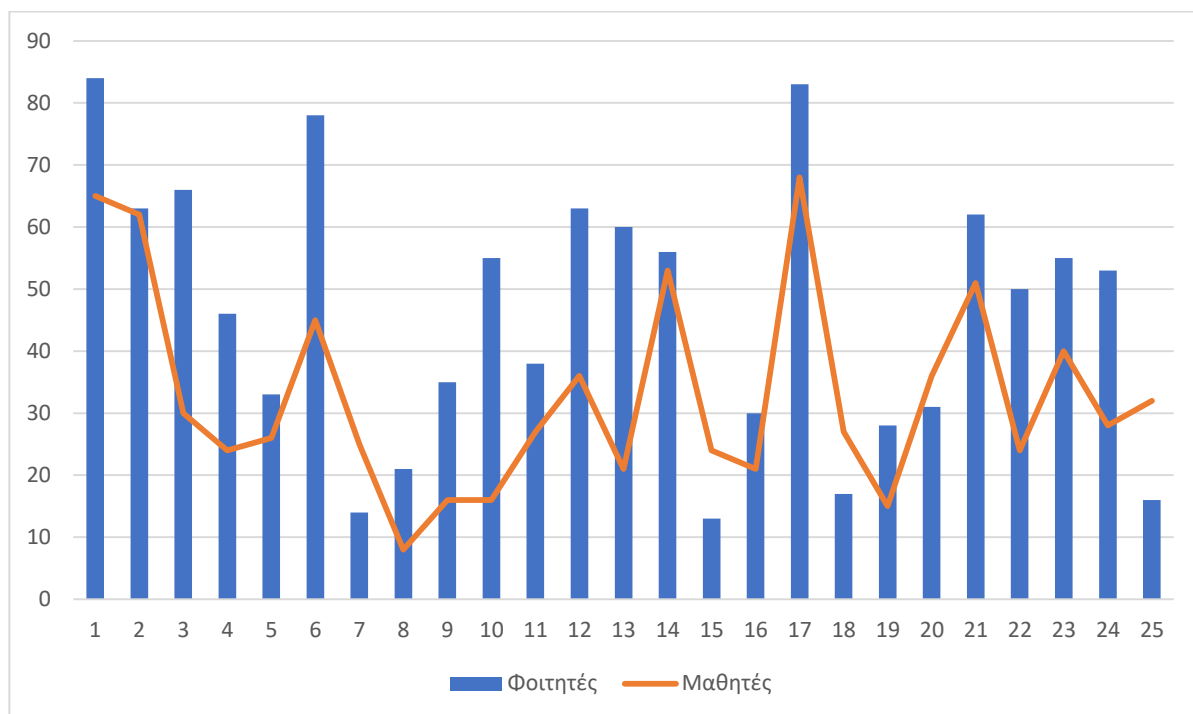
Πίνακας 2 Δείγματα

	Μαθητές	Φοιτητές
items	25	32
sample units	123	98
missing values	0	0

Στον παρακάτω πίνακα και το σχετικό συγκριτικό γράφημα (Εικόνα 6) παρουσιάζονται τα ποσοστά των σωστών απαντήσεων τόσο των μαθητών όσο και των φοιτητών στις ερωτήσεις που ήταν κοινές στα δύο ερωτηματολόγια.

Πίνακας 3 Ποσοστό σωστών απαντήσεων στις κοινές ερωτήσεις

# Ερώτηση	Ποσοστό μαθητών επι % που απάντησαν σωστά	Ποσοστό φοιτητών επι % που απάντησαν σωστά
1	65	84
2	62	63
3	30	66
4	24	46
5	26	33
6	45	78
7	25	14
8	8	21
9	16	35
10	16	55
11	27	38
12	36	63
13	21	60
14	53	56
15	24	13
16	21	30
17	68	83
18	27	17
19	15	28
20	36	31
21	51	62
22	24	50
23	40	55
24	28	53
25	32	16



Εικόνα 6 Η "μεγάλη" εικόνα

Παρατηρούμε πως στις ερωτήσεις **Q7, Q15, Q18, Q20** και **Q25** η απόδοση των μαθητών ήταν καλύτερη απ' ό τι των φοιτητών. Στο επόμενο τμήμα θα λάβουμε υπόψη παραμέτρους δυσκολίας και διάκρισης από την στατιστική ανάλυση για να ερμηνευθούν κατάλληλα τα αποτελέσματα για αυτές τις τέσσερις ερωτήσεις.

Στους δύο παρακάτω πίνακες παρουσιάζονται οι συχνότητες βαθμολογίας που εμφανίζονται τόσο στους μαθητές όσο και στους φοιτητές.

Πίνακας 4 Συχνότητα Βαθμολογίας Μαθητές

Συχνότητα συνολικής βαθμολογίας Μαθητές	
Βαθμολογία	Συχνότητα
0	0
1	2
2	2
3	1
4	2
5	9
6	22
7	16
8	20
9	11
10	13
11	11
12	5
13	6
14	1
15	0
16	0
17	0
18	0
19	1
20	0
21	1
22	0
23	0
24	0
25	0

Πίνακας 5 Συχνότητα Βαθμολογίας Φοιτητές

Συχνότητα συνολικής βαθμολογίας Φοιτητές	
Βαθμολογία	Συχνότητα
0	0
1	0
2	0
3	0
4	0
5	1
6	1
7	5
8	4
9	6
10	11
11	8
12	9
13	4
14	4
15	6
16	9
17	8
18	3
19	4
20	3
21	4
22	2
23	1
24	0
25	0
26	0
27	2
28	1
29	0
30	0
31	0
32	0

4.1.1 Μοντέλο μιας παραμέτρου

Στο σημείο αυτό γίνεται ανάλυση για τα δεδομένα μας με το μοντέλο μιας παραμέτρου τόσο για τους μαθητές όσο και για τους φοιτητές. Στο μοντέλο Rasch γνωρίζουμε πως η παράμετρος διάκρισης (Discrimn) είναι ίση με την μονάδα οπότε θεωρείται και περιττό να την συμπε-

ριλάβουμε σε διαφορετική στήλη στους δύο πίνακες που ακολουθούν. Η παράμετρος δυσκολίας (Dffclt) είναι το z-score. Η βαθμολογία z-score, γνωστή και ως τυπική βαθμολογία, αντιπροσωπεύει τον αριθμό των τυπικών αποκλίσεων (Standar Deviation) που ένα σημείο δεδομένων απέχει από τον μέσο όρο (μέσο) της ομάδας. Με απλά λόγια, δείχνει πόσες τυπικές αποκλίσεις πάνω ή κάτω από το μέσο όρο βρίσκεται μια παρατήρηση. Οι βαθμολογίες Z, επομένως, είναι ένας χρήσιμος τρόπος τυποποίησης τιμών και κινείται μεταξύ του (-3, 3), με το 3 να υποδηλώνει ότι η ερώτηση είναι δύσκολη ενώ το -3 εύκολη. Επειδή από μαθηματικής απόψεως, το 1PL και τα μοντέλα Rasch είναι ισοδύναμα και οι τιμές από ένα μοντέλο μπορούν να μετατραπούν στο άλλο με κατάλληλη επανακλιμάκωση επιλέγουμε τυχαία το μοντέλο Rasch και παρουσιάζουμε σε αυτό και την πιθανότητα $P(x=1|z=0)$ που μας δίνει τη πιθανότητα να λάβουμε μια σωστή απάντηση για το «μέσο» εξεταζόμενο.

Πίνακας 6 Αποτελέσματα 1PL (μαθητές)

Μαθητές				
#	One-parameter logistic model		Rasch	
	Dffclt	Dscrmn	Dffclt	$P(x=1 z=0)$
Q1	-1.3616137	0.535702	-0.68477291	0.66480313
Q2	-1.0867644	0.535702	-0.53022146	0.62953476
Q3	1.5508586	0.535702	0.93075241	0.28277209
Q4	2.2369482	0.535702	1.29318224	0.21531467
Q5	1.9564345	0.535702	1.15088420	0.24032762
Q6	0.2500509	0.535702	0.19962039	0.45025997
Q7	2.0387786	0.535702	1.19732069	0.23195219
Q8	4.5613102	0.535702	2.62464713	0.06756892
Q9	3.1152625	0.535702	1.79203107	0.14282389
Q10	3.1361843	0.535702	1.79276843	0.14273364
Q11	1.8721719	0.535702	1.10538494	0.24873228
Q12	1.0383906	0.535702	0.64731419	0.34359503
Q13	2.4735569	0.535702	1.44549988	0.19069510
Q14	-0.3406576	0.535702	-0.12514545	0.53124559
Q15	2.2007731	0.535702	1.29316328	0.21531787
Q16	2.4719064	0.535702	1.44550601	0.19069416
Q17	-1.6787266	0.535702	-0.84589448	0.69970521
Q18	1.8642731	0.535702	1.10544304	0.24872142
Q19	3.2035514	0.535702	1.85789351	0.13494877
Q20	1.0383280	0.535702	0.64731455	0.34359495
Q21	-0.1684249	0.535702	-0.05288515	0.51321821
Q22	2.1140550	0.535702	1.24477429	0.22360605

Q23	0.6958193	0.535702	0.45686387	0.38773006
Q24	1.8699410	0.535702	1.10540122	0.24872924
Q25	1.3995497	0.535702	0.84717502	0.30002580

Στον πρώτο πίνακα παρατηρούμε ότι οι τιμές για την παράμετρο δυσκολίας είναι στο προαναφερθέν διάστημα και η παράμετρος διάκρισης για το 1-PL είναι θετική. Συνεπώς δεν προι-
δεαζόμαστε «αρνητικά» για κάποια από τις ερωτήσεις.

Πίνακας 7 Αποτελέσματα IPL (φοιτητές)

Φοιτητές				
#	One-parameter logistic model		Rasch	
	Dffclt	Dscrmn	Dffclt	P(x=1 z=0)
Q1	-2.32629991	0.7829215	-1.86041964	0.86534585
Q2	-0.79878125	0.7829215	-0.64603502	0.65611641
Q3	-0.99188432	0.7829215	-0.79889337	0.68973771
Q4	0.22140616	0.7829215	0.16988744	0.45763000
Q5	1.03516829	0.7829215	0.82073782	0.30560706
Q6	-1.78221954	0.7829215	-1.42865152	0.80669112
Q7	2.55429033	0.7829215	2.03887411	0.11518143
Q8	1.86331983	0.7829215	1.48290586	0.18498891
Q9	0.90306336	0.7829215	0.71480050	0.32853897
Q10	-0.31126641	0.7829215	-0.25695813	0.56388839
Q11	0.71036809	0.7829215	0.56082853	0.36335578
Q12	-0.79874104	0.7829215	-0.64603108	0.65611552
Q13	-0.61262007	0.7829215	-0.49767557	0.62191293
Q14	-0.37116938	0.7829215	-0.30462515	0.57557278
Q15	2.67678489	0.7829215	2.13465735	0.10577366
Q16	1.23942835	0.7829215	0.98624925	0.27165356
Q17	-2.22428145	0.7829215	-1.78125419	0.85585166
Q18	2.23211499	0.7829215	1.78010481	0.14429019
Q19	1.30989745	0.7829215	1.04332790	0.26050838
Q20	1.17024869	0.7829215	0.93009951	0.28290453
Q21	-0.73620575	0.7829215	-0.59617669	0.64478111
Q22	-0.01741313	0.7829215	-0.02005562	0.50501374
Q23	-0.31082369	0.7829215	-0.25698660	0.56389539
Q24	-0.19216611	0.7829215	-0.16198322	0.54040749
Q25	2.33455292	0.7829215	1.86221662	0.13444490
Q26	1.86311800	0.7829215	1.48291496	0.18498753
Q27	1.69402448	0.7829215	1.34893933	0.20604383
Q28	2.13415744	0.7829215	1.70143200	0.15427833
Q29	-1.05472403	0.7829215	-0.85108679	0.70079507
Q30	-0.92574949	0.7829215	-0.74733483	0.67859770
Q31	3.08683185	0.7829215	2.46272378	0.07851305
Q32	-0.67455845	0.7829215	-0.54668217	0.63336549

Όπως έχει και πρότερα αναφερθεί και τα δύο μοντέλα 1PL και Rasch απαιτούν τα στοιχεία να έχουν σταθερή τιμή για την παράμετρο διάκρισης, για το μοντέλο Rasch, αυτή η σταθερά είναι 1,0, ενώ για το μοντέλο 1PL η σταθερά a δεν χρειάζεται να είναι ίση με 1. Μια πιο σύντομη/υπεραπλουστευμένη απάντηση είναι ότι το Rasch κεντράρει τη βαθμονόμηση σε αντικείμενα και το IRT κεντράρει τη βαθμονόμηση σε άτομα. Εάν τα στοιχεία είναι 0,5 logit πάνω από τα άτομα κατά μέσο όρο, ο Rasch θα έλεγε ότι η μέση δυσκολία αντικειμένου είναι 0,0 και η μέση ικανότητα ατόμου είναι 0,5. Το IRT θα έλεγε ότι η μέση δυσκολία αντικειμένου είναι -0,5 και η μέση ικανότητα ατόμου είναι 0,0. Λόγω αυτής της μικρής αλλαγής, μπορεί να ακούσετε ότι το μοντέλο Rasch είναι "μέτρηση χωρίς άτομα" ή τέτοιους όρους. Φυσικά, τόσο τα μοντέλα Rasch όσο και 1PL υποθέτουν ότι τα είδη δεν διαφέρουν σε ποιότητα (παράμετρος διάκρισης).

4.1.2 Μοντέλο δύο παραμέτρων

Στο μοντέλο αυτό η παράμετρος διάκρισης δεν παραμένει σταθερή οπότε λαμβάνουμε περισσότερη πληροφορία συγκριτικά με το προηγούμενο μοντέλο.

Πίνακας 8 Αποτελέσματα 2PL (μαθητές)

Μαθητές			
#	two-parameter logistic model		
	Dffclt	Dscrmn	P(x=1 z=0)
Q1	-0.77528010	0.96155168	0.67819124
Q2	-2.40787048	0.20161121	0.61903485
Q3	1.38301960	0.66612134	0.28470146
Q4	1.46063586	0.94136217	0.20181526
Q5	2.33421667	0.46905208	0.25070249
Q6	0.14773760	1.50444675	0.44466179
Q7	-1.93590247	- 0.60629908	0.23618037
Q8	-5.83620268	-0.42819524	0.07592591
Q9	36.50302952	0.04491893	0.16250900
Q10	3.15857862	0.55007730	0.14963554
Q11	1.60656085	0.68483202	0.24969787
Q12	0.81038074	0.81646746	0.34036917
Q13	5.30554217	0.25155758	0.20839135
Q14	-0.18663998	0.70472793	0.53283528
Q15	-21.79700233	-0.05398770	0.23563345
Q16	7.91930995	0.16726739	0.21004697
Q17	-0.72158988	1.50831670	0.74807768

Q18	7.07276331	0.14251344	0.26737852
Q19	-7.11171806	-0.24183813	0.15188603
Q20	0.89618260	0.72189094	0.34367806
Q21	-0.06414096	1.60192023	0.52566460
Q22	-6.62668200	- 0.17187231	0.24251445
Q23	1.08749328	0.39200404	0.39500978
Q24	2.06972814	0.51208051	0.25733479
Q25	5.88848037	0.13082227	0.31640462

Παρατηρούμε ότι η ερώτηση Q7 μιας και της αντιστοιχεί αρνητική παράμετρος διάκρισης αλλά και για τις Q8, Q9, Q13, Q15, Q16, Q18, Q19, Q22 και Q25 έχουν παράμετρο δυσκολίας εκτός των ορίων μας σημαίνει ότι δεν διακρίνουν-διαφοροποιούν καλά την υψηλή ικανότητα.

Πίνακας 9 Αποτελέσματα 2PL (φοιτητές)

Φοιτητές			
#	two-parameter logistic model		
	Dffclt	Dscrmn	P(x=1 z=0)
Q1	-1.98933065	0.95377386	0.86959375
Q2	-1.13693031	0.51171373	0.64147754
Q3	-0.56287531	2.52206623	0.80527708
Q4	0.10084029	1.43334466	0.46392806
Q5	0.84362568	0.99204002	0.30218588
Q6	-1.50630494	0.97732994	0.81338500
Q7	19.54868471	0.09180992	0.14248974
Q8	1.58122831	0.95999217	0.17976111
Q9	0.65698346	1.13773127	0.32137269
Q10	-0.30838180	0.84021648	0.56441688
Q11	0.50365888	1.15417798	0.35863076
Q12	-0.71633344	0.92681601	0.66013801
Q13	-0.94358987	0.46605399	0.60820275
Q14	-0.35711345	0.86953171	0.57701254
Q15	-1.63574439	-1.54933398	0.07348735
Q16	1.09535169	0.90356536	0.27096702
Q17	-1.49869332	1.35868429	0.88455098
Q18	2.26774353	0.76698646	0.14939824
Q19	1.51860105	0.65339386	0.27046896
Q20	2.85758781	0.29169466	0.30289642
Q21	-0.87076424	0.63785636	0.63539274
Q22	-0.05870479	1.53348325	0.52249052
Q23	-0.23778640	1.82762053	0.60696759
Q24	-0.18745453	0.94186612	0.54402496
Q25	-6.80514445	-0.24301495	0.16060252
Q26	2.24560778	0.62391580	0.19764634

Q27	1.25488587	1.15439279	0.19021231
Q28	3.77356912	0.40905100	0.17601513
Q29	-0.93689262	0.93023352	0.70506373
Q30	-0.65217720	1.39106029	0.71243050
Q31	15.47314604	0.14106799	0.10130886
Q32	-0.56578953	1.04249773	0.64332712

Προϊδεαζόμαστε «αρνητικά» για τις ερωτήσεις **Q7, Q28, Q31** λόγω συντελεστή δυσκολίας εκτός των ορίων ακόμη η **Q15** (αρνητικός συντελεστής διάκρισης), **Q25** (λόγω συντελεστή δυσκολίας και διάκρισης), μιας και οι τιμές

4.1.3 Μοντέλο τριών παραμέτρων

Πίνακας 10 Αποτελέσματα 3PL (μαθητές)

Μαθητές				
#	three-parameter logistic model			
	Gussng	Dffclt	Dscrmn	P(x=1 z=0)
Q1	5.696891e-01	0.7019398	61.30249654	0.5696891
Q2	5.801479e-01	-3.7231804	-0.66685137	0.6125068
Q3	2.568558e-01	1.6653769	316.76154761	0.2568558
Q4	1.327067e-01	1.3587914	749.31006464	0.1327067
Q5	2.437079e-01	2.0580060	215.16318406	0.2437079
Q6	2.888333e-01	0.6792073	327.59649246	0.2888333
Q7	2.495188e-216	-4.3193491	-0.25447877	0.2498931
Q8	7.438020e-02	2.7549131	429.94079518	0.0743802
Q9	1.487639e-01	2.3941670	625.55786024	0.1487639
Q10	8.544011e-20	3.3207131	0.52689517	0.1480891
Q11	2.222947e-01	1.6010444	292.00071093	0.2222947
Q12	4.289709e-177	0.8154983	0.84932740	0.3334494
Q13	1.983521e-01	2.2713135	488.63201693	0.1983521
Q14	4.696177e-01	1.3579448	189.36629654	0.4696177
Q15	6.357060e-99	15.7666240	0.07480601	0.2351532
Q16	1.983622e-01	2.3475065	554.33996304	0.1983622
Q17	1.181811e-01	-0.1546990	368.59071087	1.0000000
Q18	2.637953e-01	3.0931965	60.30240735	0.2637953
Q19	1.645292e-112	-10.6670557	-0.15992088	0.1536982
Q20	3.298923e-01	2.0444237	77.21352632	0.3298923
Q21	2.371744e-01	0.5936200	1749.92369457	0.2371744
Q22	1.973106e-01	-1.3634273	-466.50447487	0.1973106
Q23	3.884266e-01	2.3320456	136.26793236	0.3884266
Q24	2.561981e-01	2.4778606	169.77241516	0.2561981
Q25	3.170774e-01	-26.5640415	-2.55435707	0.3170774

Πίνακας 11 Αποτελέσματα 3PL (φοιτητές)

Φοιτητές				
#	three-parameter logistic model			
	Gussng	Dffclt	Dscrmn	P(x=1 z=0)
Q1	3.401912e-01	-1.49408443	0.8537360	0.85595845
Q2	5.762638e-01	1.14749843	4.9008005	0.57778854
Q3	1.272799e-05	-0.52685328	2.1855151	0.75977796
Q4	2.841169e-04	0.15508703	1.4993777	0.44228553
Q5	1.585194e-01	1.17447387	1.7564016	0.25340524
Q6	5.632442e-08	-1.34015070	1.1518676	0.82399851
Q7	1.304399e-01	3.20291534	1.8547294	0.13272135
Q8	4.899682e-02	1.47736260	1.4137556	0.15380391
Q9	1.632529e-01	0.98009495	2.1600997	0.25315768
Q10	1.855488e-01	0.23533119	1.0929019	0.54069321
Q11	7.608869e-08	0.56027666	1.1157619	0.34861444
Q12	2.888458e-01	0.06920488	1.2896332	0.62856600
Q13	5.205640e-01	1.04869262	3.3923714	0.53385313
Q14	3.339368e-01	0.59334118	1.5953652	0.52014805
Q15	6.870065e-09	-1.77868620	-1.3877394	0.07810676
Q16	1.916522e-01	1.35782646	2.5865210	0.21507122
Q17	4.421272e-07	-1.63077723	1.2015549	0.87647546
Q18	1.189712e-01	1.42925790	23.0319150	0.11897121
Q19	1.590237e-01	1.67043390	1.3605018	0.23758383
Q20	2.347282e-01	2.76520234	0.9512109	0.28616537
Q21	3.584546e-01	0.38711470	1.1808153	0.60716341
Q22	2.155665e-01	0.37637046	4.1017304	0.35361725
Q23	6.202220e-08	-0.18850456	1.5174513	0.57102800
Q24	3.741691e-01	0.67752865	24.1012574	0.37416912
Q25	1.415843e-01	-2.42434062	-2.6464214	0.14298594
Q26	1.931747e-01	2.08428785	20.3586975	0.19317473
Q27	1.034761e-01	0.82559768	19.5442908	0.10347614
Q28	2.669717e-06	3.53565128	0.4396374	0.17445325
Q29	3.910604e-01	0.12932368	1.6859571	0.66246869
Q30	3.996939e-01	0.05109669	26.0628869	0.52508211
Q31	8.460318e-02	3.37551075	1.4949465	0.09045491
Q32	1.027248e-06	-0.51634947	1.0917541	0.63731475

Από τις τιμές που λαμβάνουμε στο μοντέλο τριών παραμέτρων τόσο στους μαθητές όσο και στους φοιτητές, παρατηρούμε ότι η Παράμετρος Πρόβλεψης (Gussng) για κάθε ερώτηση είναι πολύ κοντά στο 0. Αυτό, μας προϊδεάζει ότι δεν είναι κατάλληλο μοντέλο για τα δεδομένα μιας και η παράμετρος εικασίας είναι πολύ μικρή ώστε να θεωρηθεί ως κριτήριο επιλογής

του για την ανάλυση των δεδομένων, δηλαδή δεν μπορούμε να λάβουμε υπόψη την επιπλέον πληροφορία που θεωρητικά θα μας παρείχε λόγω της εικασίας. Επιπλέον, ένας ακόμη αποτρεπτικός παράγοντας για να κάνουμε χρήση αυτού του μοντέλου είναι ότι οι παράμετροι διάκρισης τόσο στους μαθητές όσο και στους φοιτητές σε μεγάλο ποσοστό είναι είτε αρνητικές είτε αρκετά μεγάλες. Μια πιθανή αιτία είναι ίσως το μέγεθος του δείγματος και να χρειάζεται επιπλέον ποσότητα για να ανταποκριθεί με «λογικές» τιμές στις παραμέτρους.

4.2 Σύγκριση μοντέλων

4.2.1 Επιλογή μοντέλου-μαθητές

Οι πίνακες από τα μοντέλα Rasch, two-parameter logistic model (2PL) και three-parameter logistic model (3PL) μας δίνουν μια όσιμη σύγκριση για το καταλληλότερο εκ των τριών μοντέλων με το μοντέλο 2PL να αναδεικνύεται ως το καταλληλότερο. Σαφώς, δεν μπορούμε να επιλέξουμε μοντέλο στατιστικής ανάλυσης τόσο απλοϊκά. Θα προβούμε σε αυστηρότερα κριτήρια ώστε να αποφανθούμε. Επομένως, χρησιμοποιούμε Ανονα και ελέγχουμε με Likelihood Ratio Test με βάση την ύπαρξη ή μη στατιστικής σημαντικότητας.

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRTrasch	3604.29	3674.59	1777.14			
IRT2pl	3536.00	3676.61	-1718.00	118.28	25	<0.001

Μια χαμηλότερη τιμή AIC ή BIC υποδηλώνει καλύτερη εφαρμογή του μοντέλου. όπου log.Lik είναι η τιμή του λογαρίθμου της Πιθανοφάνειας.

Πιο συγκεκριμένα, το κριτήριο BIC (Bayesian Information Criterion) θα επιφέρει υψηλότερη τιμωρία για μοντέλα με περίπλοκη παραμετροποίηση σε σύγκριση με το κριτήριο AIC (Akaike's Information Criteria). Το αν αποφασίσετε να χρησιμοποιήσετε μάλλον AIC ή BIC εξαρτάται από εσάς, καθώς αυτά τα κριτήρια είναι ελαφρώς διαφορετικά, αλλά κανένα δεν είναι καλύτερο από το άλλο. Καθώς η πολυπλοκότητα του μοντέλου αυξάνεται, η τιμή BIC αυξάνεται και όσο αυξάνεται η πιθανότητα, η BIC μειώνεται. Άρα, το χαμηλότερο είναι καλύτερο. Στη συγκεκριμένη περίπτωση παρατηρούμε ότι οι τιμές BIC μεταξύ των δύο μοντέλων δεν δίνουν καλή εικόνα ώστε να λάβουμε αυτό ως κριτήριο οπότε θα χρησιμοποιήσουμε το κριτήριο AIC που έχει χαμηλότερη τιμή στο 2PL.

Επιπλέον, παρατηρούμε από τον παραπάνω πίνακα τη σύγκριση μεταξύ των μοντέλων Rasch και 2Pl ότι η πιθανοφάνεια μας δίνει σημαντικότητα δηλαδή από τις τιμές βλέπουμε βελτίωση στο Μοντέλο δύο Παραμέτρων.

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRT2pl	3536.00	3676.61	-1718.00			
IRT3pl	3537.08	3747.99	-1693.54	48.93	25	0.003

Από τις τιμές του πίνακα δεν προκύπτει βελτίωση στην προσαρμογή του μοντέλου με τις τρεις διαστάσεις οπότε παραμένει πιο ευέλικτη η επιλογή των δύο παραμέτρων.

Επιπλέον, η μονοδιάσταση (unidimensionality) είναι απαραίτητη προϋπόθεση για οποιεσδήποτε αναλύσεις που βασίζονται στην κλίμακα "συνολική βαθμολογία" και η ευρεία χρήση βαθμολογιών που αθροίζονται με αυτόν τον τρόπο σε ερευνητικές μελέτες μπορεί να αποκρύψει πιο συγκεκριμένες ομάδες στοιχείων σε κάθε υποκλίμακα. Επομένως, είναι ένα σημαντικό χαρακτηριστικό ενός οργάνου μέτρησης επειδή, με ψυχομετρικούς όρους, υποδηλώνει ότι υπάρχει ένα μόνο λανθάνον χαρακτηριστικό (π.χ. σε αυτήν την περίπτωση «μαθηματική ικανότητα») που βασίζεται στις απαντήσεις (Hattie, 1985). Η μονοδιάσταση θα πρέπει να εξετάζεται σε οποιοδήποτε εργαλείο μέτρησης για να διασφαλιστεί η ορθότητα των αξιολογήσεων που γίνονται σχετικά με τη συνολική έννοια που εξετάζεται και υποδεικνύει ότι υπάρχει μόνο ένα λανθάνον χαρακτηριστικό για τις απαντήσεις [30].

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRT2pl	3536.0	3676.61	-1718.0			
IRT2pl_2	3544.6	3755.51	-1697.3	41.41	25	0.021

4.2.2 Επιλογή Μοντέλου-φοιτητές

Ακριβώς, την ίδια σύγκριση οφείλουμε να πραγματοποιήσουμε και στα δεδομένα που συλλέχθηκαν από τους φοιτητές ώστε να συμπεράνουμε για το καταλληλότερο μοντέλο. Για να επιβεβαιώσουμε τον ισχυρισμό μας χρησιμοποιούμε Ανονα και ελέγχουμε με Likelihood Ratio Test με βάση τη στατιστική σημαντικότητα.

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRTrasch	3550.83	3633.55	-1743.42			
IRT2pl	3513.27	3678.71	-1692.64	101.56	32	<0.001

Παρατηρούμε από τον παραπάνω πίνακα τη σύγκριση μεταξύ των μοντέλων Rasch και 2PI ότι η πιθανοφάνεια μας δίνει σημαντικότητα δηλαδή από τις τιμές βλέπουμε βελτίωση στο Μοντέλο δύο Παραμέτρων. Επιπλέον η τιμή AIC στο Rasch Model είναι μεγαλύτερη.

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRT2pl	3513.27	3678.71	-1692.64			
IRT3pl	3519.82	3767.98	-1663.91	57.45	32	0.004

Παρατηρούμε από τον παραπάνω πίνακα τη σύγκριση μεταξύ των μοντέλων 2PI και 3PI ότι η πιθανοφάνεια δεν μας δίνει στατιστική σημαντικότητα υπέρ του 3PL οπότε επιλέγουμε να αναλύσουμε τα δεδομένα με το Μοντέλο 2 παραμέτρων. Επιπλέον η τιμή AIC είναι μικρότερη στο IRT2pl.

Ελέγχουμε με Ανονα σχετικά με το το 2pl αν έχει παραβιαστεί η μέτρηση οπότε δημιουργούμε με το 2pl_2 παραγόντων και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Η μονοδιάσταση θα πρέπει να εξετάζεται σε οποιοδήποτε εργαλείο μέτρησης για να διασφαλιστεί η ορθότητα των αξιολογήσεων που γίνονται σχετικά με τη συνολική έννοια που εξετάζεται και υποδεικνύει ότι υπάρχει μόνο ένα λανθάνον χαρακτηριστικό για τις απαντήσεις.

Likelihood Ratio Table						
	AIC	BIC	log.Lik	LRT	df	p.value
IRT2pl	3513.27	3678.71	-1692.64			
IRT2pl_2	3518.71	3766.87	-1663.36	58.56	32	0.003

Λαμβάνοντας, υπόψη την παραπάνω σύγκριση, είμαστε σε θέση να ανάγουμε το two-parameter logistic model (2PL) ως το καταλληλότερο μοντέλο. Οπότε, ενδιαφερόμαστε, και για περαιτέρω πληροφορίες σχετικές με τις τιμές που προκύπτουν για τις παραμέτρους όπως είναι τα 95% Διαστήματα Εμπιστοσύνης για κάθε παράμετρο κάθε ερώτησης ξεχωριστά, και παρακάτω πίνακα ως ακολούθως.

Πίνακας 12 Επιπλέον πληροφορία από το επιλεγθέν μοντέλο

Μαθητές	
log.Lik	-3420.066 Integration method: Gauss-Hermite
Dffc1t	Optimization: max(grad): 0.0057 quasi-Newton: BFGS

	convergence:0				
#	Coef	Std.error	z.vals	95% Confint	
Q1	-0.77528010	0.3078	-2.5186	-1.378568	-0.171992
Q2	-2.40787048	2.9716	-0.8103	-8.232206	3.4164655
Q3	1.38301960	0.5957	2.3218	0.2154476	2.5505916
Q4	1.46063586	0.4791	3.0487	0.5215999	2.3996719
Q5	2.33421667	1.3308	1.7540	-0.274151	4.9425847
Q6	0.14773760	0.1737	0.8506	-0.192714	0.4881896
Q7	-1.93590247	0.9558	-2.0254	-3.80927	-0.062534
Q8	-5.83620268	5.9565	-0.9798	-17.51094	5.8385373
Q9	36.50302952	248.3185	0.1470	-450.2012	523.20729
Q10	3.15857862	1.6691	1.8924	-0.112857	6.4300146
Q11	1.60656085	0.6644	2.4181	0.3043369	2.9087849
Q12	0.81038074	0.3570	2.2700	0.1106607	1.5101007
Q13	5.30554217	5.6900	0.9324	-5.846858	16.457942
Q14	-0.18663998	0.2900	-0.6436	-0.75504	0.38176
Q15	-21.79700233	107.7581	-0.2023	-233.0029	189.40887
Q16	7.91930995	12.7230	0.6224	-17.01777	32.85639
Q17	-0.72158988	0.2237	-3.2257	-1.160042	-0.283138
Q18	7.07276331	12.5969	0.5615	-17.61716	31.762687
Q19	-7.11171806	9.2384	-0.7698	-25.21898	10.995546
Q20	0.89618260	0.4200	2.1339	0.0729826	1.7193826
Q21	-0.06414096	0.1635	-0.3923	-0.384601	0.256319
Q22	-6.62668200	10.4752	-0.6326	-27.15807	13.90471
Q23	1.08749328	0.8027	1.3548	-0.485799	2.6607853
Q24	2.06972814	1.0984	1.8843	-0.083136	4.2225921
Q25	5.88848037	10.9730	0.5366	-15.6186	27.39556
Dscrmn					
#	Coef	std.err	z.vals	95% ConfInt	
Q1	0.96155168	0.3513	2.7371	0.2730037	1.6500997
Q2	0.20161121	0.2412	0.8360	-0.271141	0.6743632
Q3	0.66612134	0.2839	2.3463	0.1096773	1.2225653
Q4	0.94136217	0.3431	2.7439	0.2688862	1.6138382
Q5	0.46905208	0.2746	1.7080	-0.069164	1.0072681
Q6	1.50444675	0.4856	3.0979	0.5526708	2.4562228
Q7	-0.60629908	0.3203	-1.8928	-1.234087	0.0214889
Q8	-0.42819524	0.4599	-0.9310	-1.329599	0.4732088
Q9	0.04491893	0.3057	0.1469	-0.554253	0.6440909
Q10	0.55007730	0.3128	1.7585	-0.063011	1.1631653
Q11	0.68483202	0.2931	2.3365	0.110356	1.259308
Q12	0.81646746	0.3027	2.6975	0.2231755	1.4097595
Q13	0.25155758	0.2738	0.9189	-0.28509	0.7882056
Q14	0.70472793	0.2861	2.4631	0.1439719	1.2654839
Q15	-0.05398770	0.2671	-0.2021	-0.577504	0.4695283
Q16	0.16726739	0.2705	0.6183	-0.362913	0.6974474
Q17	1.50831670	0.5357	2.8154	0.4583447	2.5582887
Q18	0.14251344	0.2545	0.5600	-0.356307	0.6413334

Q19	-0.24183813	0.3195	-0.7570	-0.868058	0.3843819
Q20	0.72189094	0.2929	2.4648	0.1478069	1.2959749
Q21	1.60192023	0.5372	2.9822	0.5490082	2.6548322
Q22	-0.17187231	0.2734	-0.6286	-0.707736	0.3639917
Q23	0.39200404	0.2450	1.6002	-0.088196	0.872204
Q24	0.51208051	0.2789	1.8358	-0.034563	1.0587245
Q25	0.13082227	0.2435	0.5373	-0.346438	0.6080823

Πίνακας 13 Επιπλέον πληροφορία από το επιλεγθέν μοντέλο

Φοιτητές					
	log.Lik	-1692.636	Integration method: Gauss-Hermite Optimization: max(grad): 0.003 quasi-Newton: BFGS convergence:0		
Dffelt					
#	Coef	std.err	z.vals	95% ConfInt	
Q1	-1.98933065	0.7777	-2.5581	-3.513623	-0.465039
Q2	-1.13693031	0.6769	-1.6796	-2.463654	0.1897937
Q3	-0.56287531	0.1620	-3.4742	-0.880395	-0.245355
Q4	0.10084029	0.2015	0.5005	-0.2941	0.4957803
Q5	0.84362568	0.3374	2.5005	0.1823217	1.5049297
Q6	-1.50630494	0.5270	-2.8581	-2.539225	-0.473385
Q7	19.54868471	65.5958	0.2980	-109.0191	148.11645
Q8	1.58122831	0.5043	3.1354	0.5928003	2.5696563
Q9	0.65698346	0.2824	2.3268	0.1034795	1.2104875
Q10	-0.30838180	0.2865	-1.0764	-0.869922	0.2531582
Q11	0.50365888	0.2602	1.9355	-0.006333	1.0136509
Q12	-0.71633344	0.3175	-2.2560	-1.338633	-0.094033
Q13	-0.94358987	0.6551	-1.4403	-2.227586	0.3404061
Q14	-0.35711345	0.2823	-1.2652	-0.910421	0.1961946
Q15	-1.63574439	0.4570	-3.5796	-2.531464	-0.740024
Q16	1.09535169	0.4103	2.6699	0.2911637	1.8995397
Q17	-1.49869332	0.4370	-3.4293	-2.355213	-0.642173
Q18	2.26774353	0.8315	2.7273	0.6380035	3.8974835
Q19	1.51860105	0.6591	2.3040	0.2267651	2.8104371
Q20	2.85758781	2.3894	1.1959	-1.825636	7.5408118

Q21	-0.87076424	0.4750	-1.8333	-1.801764	0.0602358
Q22	-0.05870479	0.1891	-0.3104	-0.429341	0.3119312
Q23	-0.23778640	0.1731	-1.3735	-0.577062	0.1014896
Q24	-0.18745453	0.2567	-0.7304	-0.690587	0.3156775
Q25	-6.80514445	8.9581	-0.7597	-24.36302	10.752732
Q26	2.24560778	0.9704	2.3141	0.3436238	4.1475918
Q27	1.25488587	0.3732	3.3621	0.5234139	1.9863579
Q28	3.77356912	2.4829	1.5198	-1.092915	8.6400531
Q29	-0.93689262	0.3629	-2.5813	-1.648177	-0.225609
Q30	-0.65217720	0.2269	-2.8746	-1.096901	-0.207453
Q31	15.47314604	38.2794	0.4042	-59.55448	90.50077
Q32	-0.56578953	0.2665	-2.1228	-1.08813	-0.04345
Dscrmn					
#	Coef	std.err	z.vals	95% ConfInt	
Q1	0.95377386	0.4483	2.1276	0.0751059	1.8324419
Q2	0.51171373	0.2690	1.9024	-0.015526	1.0389537
Q3	2.52206623	0.8188	3.0804	0.9172182	4.1269142
Q4	1.43334466	0.4134	3.4671	0.6230807	2.2436087
Q5	0.99204002	0.3137	3.1629	0.377188	1.606892
Q6	0.97732994	0.3975	2.4586	0.1982299	1.7564299
Q7	0.09180992	0.3088	0.2974	-0.513438	0.6970579
Q8	0.95999217	0.3228	2.9740	0.3273042	1.5926802
Q9	1.13773127	0.3417	3.3294	0.4679993	1.8074633
Q10	0.84021648	0.3020	2.7825	0.2482965	1.4321365
Q11	1.15417798	0.3512	3.2866	0.465826	1.84253
Q12	0.92681601	0.3344	2.7719	0.271392	1.58224
Q13	0.46605399	0.2587	1.8015	-0.040998	0.973106
Q14	0.86953171	0.3098	2.8071	0.2623237	1.4767397
Q15	-1.54933398	0.6282	-2.4664	-2.780606	-0.318062
Q16	0.90356536	0.3009	3.0029	0.3138014	1.4933294
Q17	1.35868429	0.5311	2.5582	0.3177283	2.3996403
Q18	0.76698646	0.3051	2.5136	0.1689905	1.3649825
Q19	0.65339386	0.2684	2.4342	0.1273299	1.1794579
Q20	0.29169466	0.2393	1.2189	-0.177333	0.7607227
Q21	0.63785636	0.2854	2.2352	0.0784724	1.1972404
Q22	1.53348325	0.4300	3.5660	0.6906833	2.3762833
Q23	1.82762053	0.5255	3.4776	0.7976405	2.8576005
Q24	0.94186612	0.3160	2.9802	0.3225061	1.5612261
Q25	-0.24301495	0.3249	-0.7480	-0.879819	0.3937891
Q26	0.62391580	0.2787	2.2387	0.0776638	1.1701678
Q27	1.15439279	0.3442	3.3543	0.4797608	1.8290248
Q28	0.40905100	0.2776	1.4738	-0.135045	0.953147
Q29	0.93023352	0.3480	2.6734	0.2481535	1.6123135
Q30	1.39106029	0.4356	3.1931	0.5372843	2.2448363
Q31	0.14106799	0.3509	0.4020	-0.546696	0.828832
Q32	1.04249773	0.3506	2.9738	0.3553217	1.7296737

Παρατηρούμε ότι και στους δύο πίνακες υπάρχουν αρνητικές τιμές της παραμέτρου διάκρισης δηλαδή, η πιθανότητα σωστής απόκρισης δεν αυξάνεται καθώς αυξάνεται το επίπεδο ικανότητας. Σε αυτά τα ερωτήματα, η πιθανότητα σωστής απόκρισης μειώνεται καθώς το επίπεδο ικανότητας αυξάνεται από χαμηλό σε υψηλό. Μερικές φορές η σωστή απάντηση σε ένα στοιχείο θα αποφέρει έναν αρνητικό δείκτη διάκρισης. Αυτό μας λέει ότι κάτι δεν πάει καλά με το ερώτημα: Είτε είναι κακώς γραμμένο είτε υπάρχει κάποια παραπληροφόρηση που επικρατεί μεταξύ των μαθητών υψηλής ικανότητας. Σε κάθε περίπτωση (Baker, Kim 2004 & Baker, 2004), είναι μια προειδοποίηση ότι το ερώτημα χρειάζεται κάποια προσοχή ή απλούστερα εξώθηση από το ερωτηματολόγιο.

Παρότι έχουμε αποφανθεί ότι το μοντέλο δύο παραμέτρων είναι το καταλληλότερο, επιχειρούμε να ερμηνεύσουμε τις «περίεργες» τιμές των παραμέτρων (τις αρνητικές ή κοντά στο μηδέν τιμές για την διακριτική ισχύ αλλά και τις εκτός ορίων (-3,3) τιμές για την παράμετρο δυσκολίας) μήπως οφείλονται σε κακή προσαρμογή. Ένα καλό κριτήριο για το πόσο καλά προσαρμόζονται τα items στο μοντέλο 2 παραμέτρων είναι το χ^2 μιας και ικανοποιούνται οι προϋποθέσεις του, και πράγματι προκύπτουν καλές τιμές για όλες τις ερωτήσεις. Στον παρακάτω πίνακα γίνεται στατιστικός έλεγχος για το πόσο καλά προσαρμόζονται τα items στο επιλεγθέν μοντέλο.

Πίνακας 14 Αποτελέσματα χ^2 test

Μαθητές		
#	X^2	$Pr(>X^2)$
Q1	13.0051	0.1117
Q2	9.3930	0.3102
Q3	19.7884	0.0112
Q4	18.1325	0.0203
Q5	6.7650	0.5622
Q6	10.3478	0.2415
Q7	5.2381	0.7318
Q8	12.3382	0.1367
Q9	9.6673	0.2892
Q10	5.3329	0.7215
Q11	9.3981	0.3098
Q12	8.9076	0.3502
Q13	4.6828	0.7909
Q14	10.1122	0.2572
Q15	19.5995	0.012
Q16	6.5048	0.5909
Q17	6.7219	0.5669
Q18	17.6838	0.0237
Q19	16.3906	0.0371
Q20	8.8449	0.3555
Q21	11.5349	0.1732
Q22	19.1305	0.0142
Q23	3.9770	0.8592
Q24	11.6355	0.1682
Q25	7.7973	0.4535

Πίνακας 15 Αποτελέσματα χ^2 test

Φοιτητές		
#	X^2	$Pr(>X^2)$
Q1	10.1801	0.2526
Q2	17.0008	0.0301
Q3	4.7434	0.7846
Q4	5.5878	0.6933
Q5	13.1468	0.1069
Q6	8.6217	0.3752
Q7	8.9770	0.3442
Q8	11.0309	0.2
Q9	5.0943	0.7474

Q10	5.4791	0.7054
Q11	16.3316	0.0379
Q12	11.1340	0.1942
Q13	7.5500	0.4786
Q14	8.3262	0.4023
Q15	7.4823	0.4856
Q16	6.8702	0.5507
Q17	3.9260	0.8637
Q18	8.7459	0.3642
Q19	5.0014	0.7574
Q20	2.5103	0.9612
Q21	7.5773	0.4758
Q22	9.2206	0.324
Q23	2.5739	0.9582
Q24	8.8866	0.352
Q25	7.0863	0.5273
Q26	7.9690	0.4365
Q27	20.8946	0.0074
Q28	4.2508	0.8338
Q29	10.3367	0.2422
Q30	8.8321	0.3567
Q31	8.8388	0.3561
Q32	9.9162	0.271

προκύψουν στατιστικά σημαντικές τιμές δηλαδή $p < 0.01$ σημαίνει ότι το item δεν έχει καλό ταίριασμα. Από τους δύο παραπάνω πίνακες που προηγήθηκαν, όλα τα item (ερωτήματα) έχουν καλή προσαρμογή fit στο μοντέλο αφού $Pr > 0.01$ για κάθε Q_i , οπότε κάθε ερώτηση που αφού ικανοποιεί το συγκεκριμένο κριτήριο αν δεν έχει τις επιθυμητές τιμές είτε στην δυσκολία είτε στην διακριτική ισχύ μπορεί το δίχως άλλο να εξωθηθεί από το ερωτηματολόγιο.

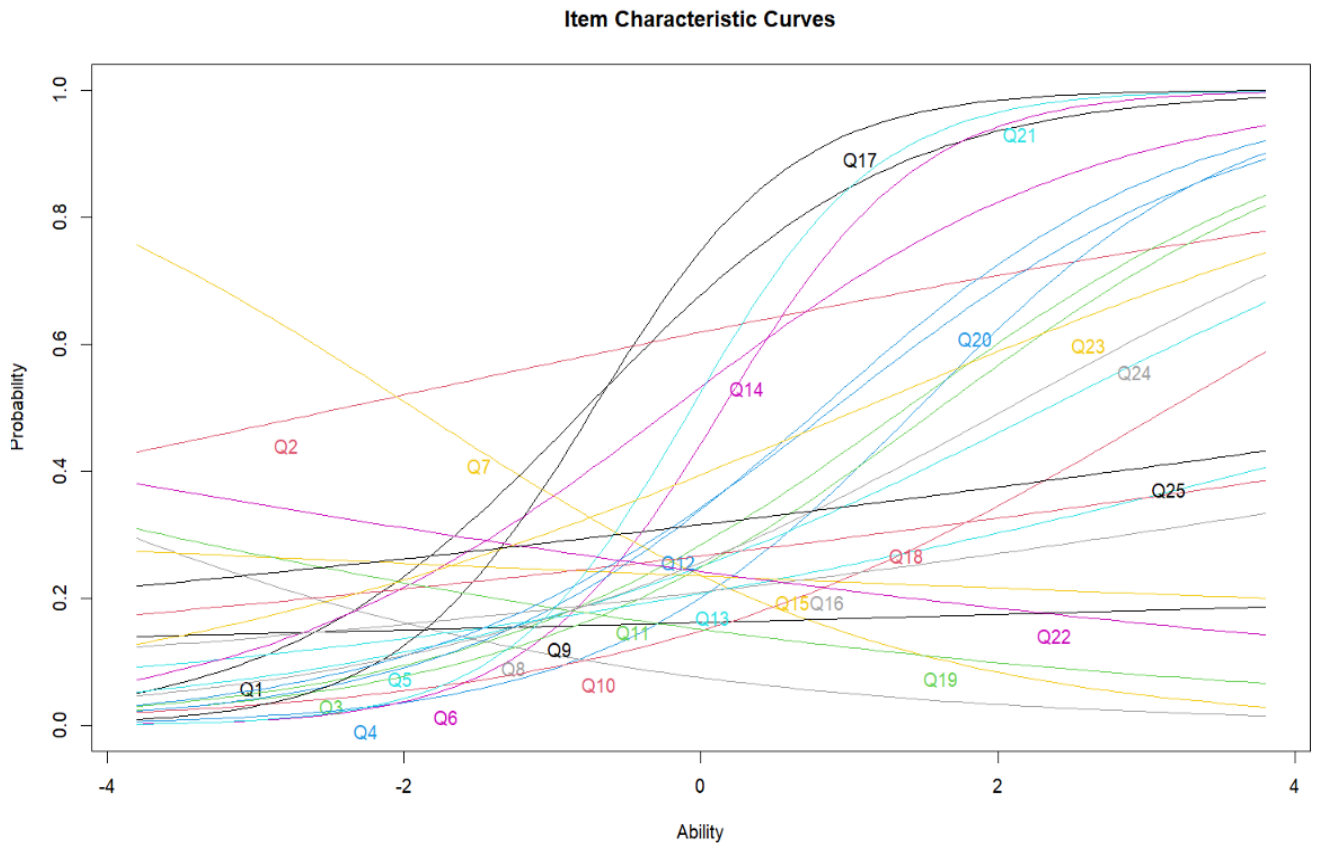
Πιο συγκεκριμένα, λαμβάνοντας υπόψη το κριτήριο (-3,3) για την δυσκολία και αποδεχόμενοι τιμές μεγαλύτερες ή ίσες του 0,3 μιας και (Baker, 2004) τιμές για τον συντελεστή διαχωρισμού μικρότερες του 0,34 θεωρούνται πολύ χαμηλές ως προς την διακριτική ισχύ καταλήγουμε να εξωθήσουμε στους

- Μαθητές, τις ερωτήσεις
 - Με βάση τον συντελεστή δυσκολίας 8, 9, 13, 15, 16, 18, 19, 22,
 - Με βάση τον συντελεστή διαχωρισμού: 2, 7, 8, 9, 13, 15, 16, 18, 19, 22
- Φοιτητές, τις ερωτήσεις
 - Με βάση τον συντελεστή δυσκολίας 7, 25, 28, 31
 - Με βάση τον συντελεστή διαχωρισμού: 7, 15, 20, 25, 31

4.3 Γραφικές

4.3.1 Χαρακτηριστικές Καμπύλες Ερωτημάτων

Στην παρακάτω εικόνα παρουσιάζεται η Χαρακτηριστική Καμπύλη για κάθε ερώτηση. Συμβάλλει στο να κατανοήσουμε αν για την χαμηλή ικανότητα η πιθανότητα σωστής απάντησης είναι μικρή και για την υψηλή ικανότητα η πιθανότητα απάντησης είναι μεγάλη με βάση κάθε S-καμπύλη.



Εικόνα 7 Χαρακτηριστικές Καμπύλες Ερωτημάτων (Μαθητές)

Επιπλέον, οι ερωτήσεις με αρνητικό συντελεστή διαχωρισμού (ισχύς διάκρισης ικανότητας), έχουν δηλαδή αρνητική κλίση και αυτό σημαίνει ότι δεν είναι καλές ερωτήσεις μιας και οι καλοί απαντούν με μεγαλύτερη συχνότητα λάθος. Έχει ενδιαφέρον να μελετηθεί, πριν απλά «αφαιρεθούν» οι ερωτήσεις με αρνητική κλίση, πως αποκρίθηκαν οι μαθητές με τις καλύτερες βαθμολογίες σε αυτές τις ερωτήσεις.

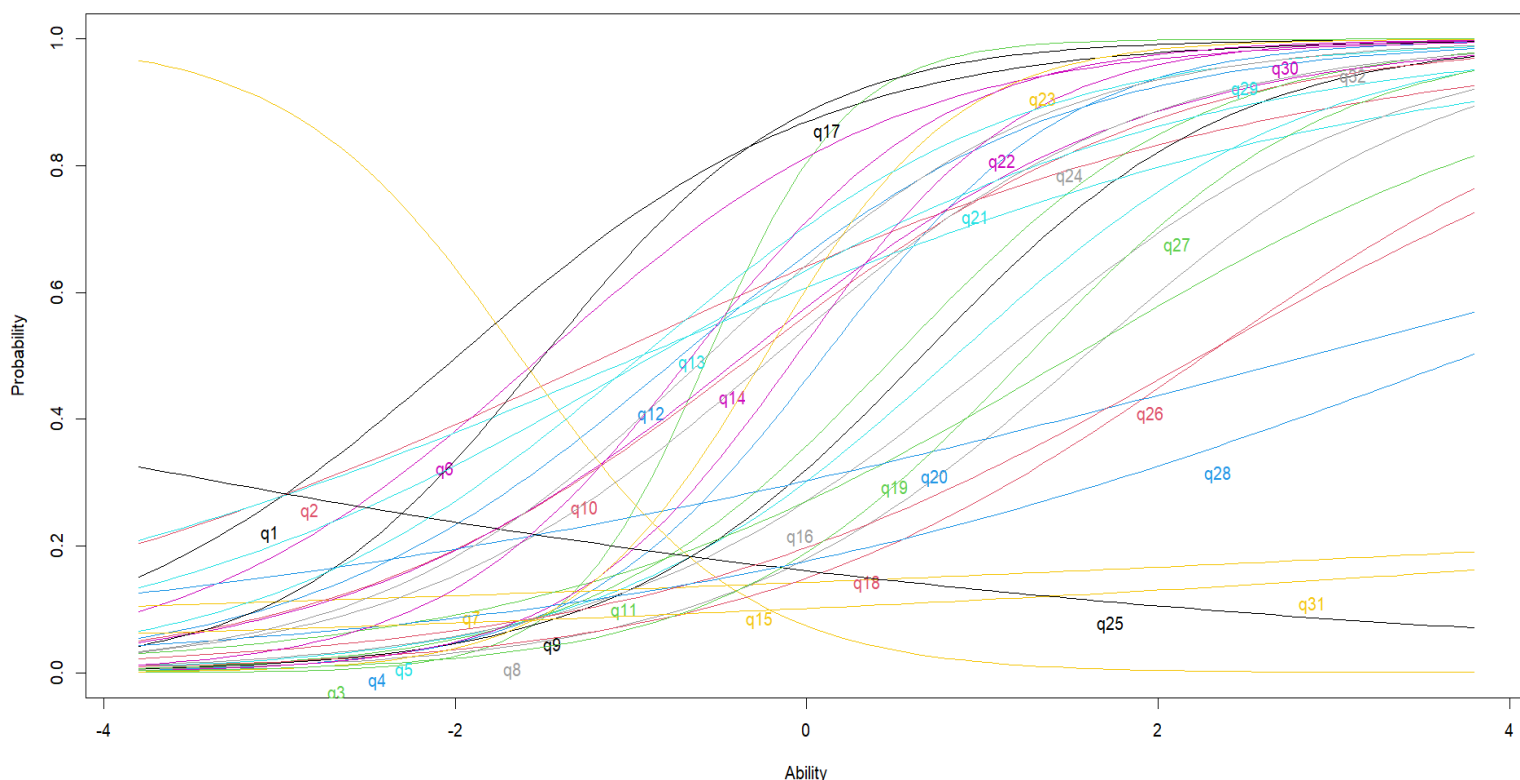
Πίνακας 16 Σωστές απαντήσεις στις ερωτήσεις με αρνητικό discrimination

Ερώτηση	Συντελεστής Διαχωρισμού	Πλήθος Σωστών Απαντήσεων από μαθητές με συνολικό σκορ μεγαλύτερο από 15/25
7	- 0.60629908	2 στους 3
8	-0.42819524	1 στους 3
15	-0.05398770	0 στους 3
19	-0.24183813	1 στους 3
22	- 0.17187231	1 στους 3

Όπως παρατηρούμε οι τρεις συνολικά σε πλήθος με βαθμολογία μεγαλύτερη από 15/30 δηλαδή οι καλοί για τα δεδομένα μας, έχουν πολύ μικρό ποσοστό κι ίσως μπορούμε να αποφανθούμε πως σε αυτές τις ερωτήσεις κάπου μπερδεύονται.

Επομένως, το γράφημα (Εικόνα 5) επιβεβαιώνει στην ουσία όσα είχαμε συμπεράνει και στην προηγούμενη παράγραφο για τις ερωτήσεις Q2, Q7, Q8, Q9, Q13, Q 15, Q 16, Q 18, Q19, Q22. Οφείλουμε να επισημάνουμε ότι και οι ερωτήσεις στις οποίες αντιστοιχούν οριζόντιες γραμμές σε αυτό το γράφημα σημαίνει ότι ανεξαρτήτως ικανότητας, η πιθανότητα σωστής απάντησης είναι η ίδια, επομένως, αυτές οι ερωτήσεις δεν είναι ιδιαίτερα «καλές» .

Item Characteristic Curves



Εικόνα 8 Χαρακτηριστικές Καμπύλες Ερωτημάτων (Φοιτητές)

Οι Χαρακτηριστικές Καμπύλες των Ερωτημάτων (Εικόνα 6) επιβεβαιώνει όσα είχαμε συμπεράνει και στην προηγούμενη παράγραφο για τις ερωτήσεις Q7, Q15, Q20, Q25, Q28, Q31 Δηλαδή καθώς αυξάνεται η ικανότητα, η πιθανότητα να απαντηθούν σωστά μειώνεται. Έχει ενδιαφέρον να μελετηθεί, πριν απλά «αφαιρεθούν» οι ερωτήσεις με αρνητική κλίση, πως αποκρίθηκαν οι μαθητές με τις καλύτερες βαθμολογίες σε αυτές τις ερωτήσεις.

Πίνακας 17 Σωστές απαντήσεις στις ερωτήσεις με αρνητικό discrimination

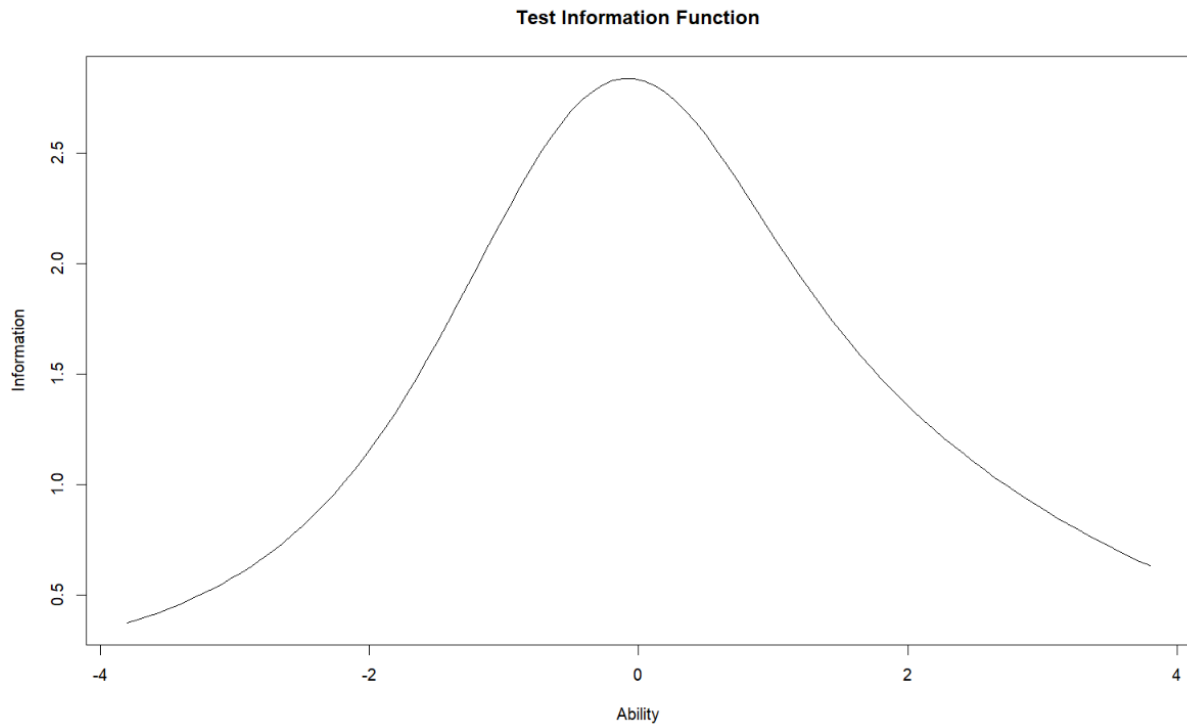
Ερώτηση	Συντελεστής Διαχωρισμού	Πλήθος Σωστών Απαντήσεων από φοιτητές με συνολικό σκορ μεγαλύτερο από 25/32
15	-1.54933398	0 στους 4
25	-0.24301495	2 στους 4

Όπως παρατηρούμε οι τέσσερις φοιτητές συνολικά σε πλήθος με βαθμολογία μεγαλύτερη από 25/32 δηλαδή οι καλοί για τα δεδομένα μας, έχουν πολύ μικρό ποσοστό κι ίσως μπορούμε να αποφανθούμε πως σε αυτές τις ερωτήσεις κάπου μερδεύονται.

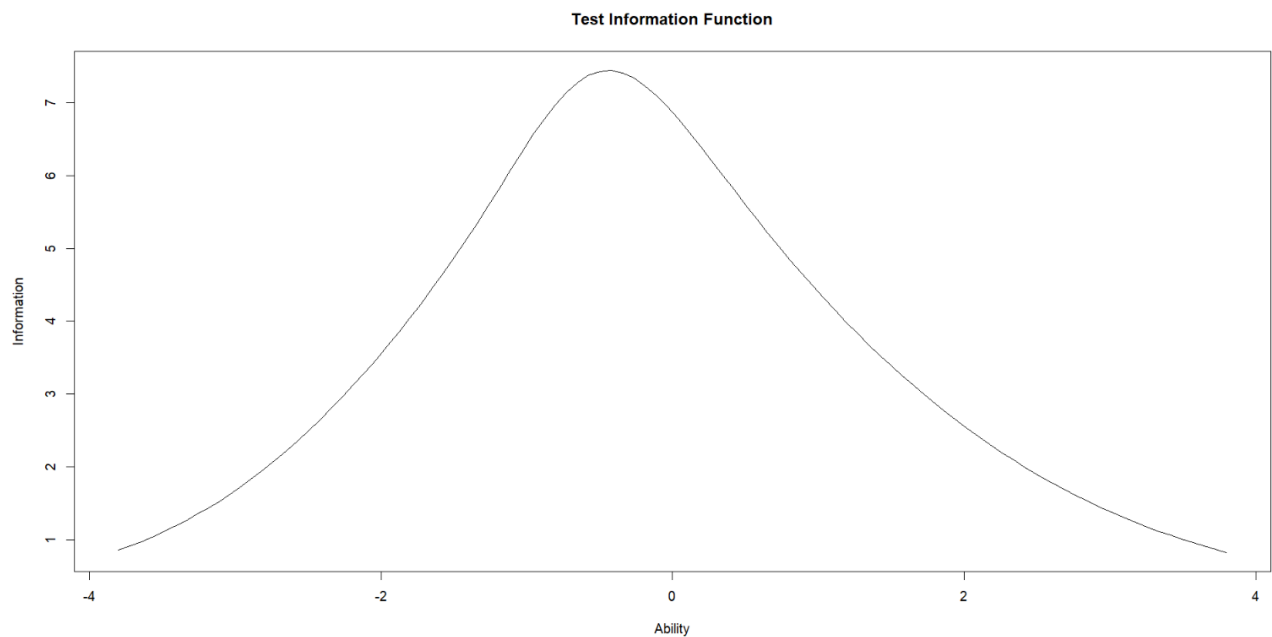
4.3.2 Συνάρτηση Πληροφορίας Ερωτήματος

Η αξιοπιστία του ερωτήματος (item) αντικαθίσταται από την πληροφορία του item στο IRT. Κάθε Συνάρτηση Απόκρισης Ερωτήματος (Item Response Function-IRF) μπορεί να μετατραπεί σε μια Συνάρτηση Πληροφορίας Ερωτήματος (Item Information Function-IIF). Οι πληροφορίες είναι ένας δείκτης που αντιπροσωπεύει την ικανότητα διαφοροποίησης των ερωτημάτων μεταξύ των ατόμων. Επιπλέον, το τυπικό σφάλμα μέτρησης είναι η αμοιβαία πληροφορία και επομένως περισσότερες πληροφορίες λιγότερα σφάλματα στη μέτρηση. Να σημειωθεί ότι μεγάλη διάκριση - ψηλά και στενά IIF, υψηλή ακρίβεια/στενό εύρος και χαμηλή διάκριση - σύντομες και ευρείες IIF. χαμηλή ακρίβεια/ευρύ φάσμα.

Για κάθε ικανότητα εκτιμά την ποσότητα της πληροφορίας που λαμβάνουμε και αυτό που θέλουμε να δούμε είναι μια κανονική καμπύλη. Δεν θέλουμε δηλαδή να δούμε πολλή κύρτωση (δηλαδή αρνητική ή θετική) ή λοξότητα (δηλαδή αρνητική ή θετική ασυμμετρία). Με απλούστερα λόγια, θέλουμε η «εικόνα» που θα μας δώσει να είμαστε σε θέση, να μπορούμε να κρίνουμε το τεστ στο σύνολό του και να δούμε σε ποιο τμήμα του εύρους χαρακτηριστικών λειτουργεί καλύτερα.



Εικόνα 9 Συνάρτηση Πληροφορίας (Μαθητές)



Εικόνα 10 Συνάρτηση Πληροφορίας (Φοιτητές)

Στην πρώτη κατανομή (Εικόνα 9) βλέπουμε ότι υπάρχει συμμετρία, οπότε λαμβάνουμε μια σχετικά καλή «ποσότητα» πληροφορίας κατά μήκος όλων των επιπέδων ικανότητας των μαθητών, δηλαδή το τεστ μας δίνει πολύ πληροφορία για τους περισσότερους μαθητές αλλά λί-

γη για τους εξαιρετικά καλούς (υψηλή ικανότητα) ή τους πολύ κακούς (χαμηλή ικανότητα). Αντίστοιχα, τα ίδια ισχύουν για τους φοιτητές (Εικόνα 10).

4.4 Συμπεράσματα

Με βάση τα όσα αναλύθηκαν παραπάνω το καταλληλότερο μοντέλο και για τις δύο ομάδες παρατήρησης (φοιτητές και μαθητές) αποδεικνύεται το μοντέλο δύο παραμέτρων εξαιτίας του ότι δίνει περισσότερη πληροφορία από τα μοντέλα μιας παραμέτρου και το μοντέλο των τριών παραμέτρων έδινε πολύ χαμηλές τιμές στην παράμετρο εικασίας οπότε δεν είναι τόσο αξιοποιήσιμα τα αποτελέσματα και επιπροσθέτως αποδεικνύεται ότι προσαρμόζονται αρκετά καλά τα ερωτήματα σε αυτό το μοντέλο. Σε κάθε φόρμα δοκιμής (ερωτηματολόγιο) με βάση τις τιμές των παραμέτρων δυσκολίας και διαχωρισμού κρίνεται αναγκαία η εξώθηση συγκεκριμένων ερωτημάτων καθώς δεν προσφέρουν καλή ισχύ διάκρισης της υψηλής και χαμηλής ικανότητας ώστε να καθίσταται έγκυρο και αξιόπιστο το σταθμισμένο εργαλείο αξιολόγησης των μαθητών επί της ικανότητας στα Μαθηματικά και της γνώσης των φοιτητών στα Μαθηματικά που πρόκειται να διδάξουν ως μελλοντικοί δάσκαλοι.

Τέλος, η σύγκριση στις κοινές ερωτήσεις απέδειξαν πως σε κάποιες αρκετά δύσκολες, δηλαδή με συντελεστή (παράμετρο) δυσκολίας αρκετά μεγάλο οι μαθητές απαντούν καλύτερα χωρίς να σημαίνει πως οι φοιτητές του τμήματος δεν είναι γνώστες αλλά σαφώς κύριο παράγοντα παίζει πάντα κατά πόσο πρόσφατες είναι οι γνώσεις στις οποίες εξετάζεσαι.

4.4.1 Μελλοντική έρευνα

Αναφορικά με μελλοντική εργασία επί του παρόντος θέματος, μια ενδιαφέρουσα ιδέα είναι η συμπερίληψη και σύγκριση με μοντέλα που χειρίζονται πολυτομικά δεδομένα, όπως το μοντέλο διαβαθμισμένης απόκρισης και το μοντέλο μερικής πίστωσης. Αυτά τα μοντέλα προβλέπουν την αναμενόμενη βαθμολογία για κάθε κατηγορία απόκρισης. Από την άλλη πλευρά, άλλα μοντέλα IRT, όπως τα μοντέλα ονομαστικής απόκρισης, προβλέπουν τις αναμενόμενες βαθμολογίες ατόμων που απαντούν σε στοιχεία με κατηγορίες μη ταξινομημένων απαντήσεων (π.χ. Ναι, Όχι, Ίσως).

Στην ουσία, εστίασαμε σε μονοδιάστατα μοντέλα IRT, που αφορούν τη μέτρηση ενός λανθάνοντος χαρακτηριστικού, ωστόσο αυτά τα μοντέλα δεν θα ήταν κατάλληλα για τη μέτρηση περισσότερων από ένα λανθάνον κατασκευών ή χαρακτηριστικών. Στην τελευταία περίπτωση, συνιστάται η χρήση πολυδιάστατων μοντέλων IRT.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Ado Abdu Bichi, Rahimah Embong, Mustafa Mamat, Danjuma A. Maiwada, (2015). *Comparison of Classical Test Theory and Item Response Theory: A Review of Empirical Studies 1 Australian Journal of Basic and Applied Sciences*, Pages: 549-556
- AERA, APA, & NCME (1999). *Improving the Likert Scale of the Children's Social Skills Test by Means of Rasch Model, The Standards for Educational and Psychological Testing*. Washington: AERA Publications Sales.
- Baker F, Kim SH (2004). *Item Response Theory*. Marcel Dekker, New York, 2nd edition.
- Baker, Frank (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD
- Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker. Available here and here
- Bartholomew D, Steele F, Moustaki I, Galbraith J (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Chapman & Hall, London.
- Bowling, A. (1995) *Research Methods in Health. Investigating Health and Health Services* (Buckingham: Open University Press).
- Cohen, L. and Manion, L. (1998). *Research Methods in Education* (London: Routledge).
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Orlando, FL: Holt, Rinehart and Winston
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. New York: Oxford University Press, Inc.
- Gehlbach, Hunter & Brinkworth, Maureen & Juraschek, Anna. (2012). *Changes in teacher-student relationships. The British journal of educational psychology*. 82. 690-704. 10.1111/j.2044-8279.2011.02058.x.

- Hattie, J. (1985). *Methodology review: Assessing unidimensionality of tests and items*. *Applied Psychological Measurement*, 9(2), 139–164. <https://doi.org/10.1177/014662168500900204>
- Kaplan Robert M., Dennis P. *Psychological Testing: Principles, Applications, and Issues*
- Li Y, Baron J. (2012). *Behavioral Research Data Analysis with R: Springer New York; Chapter 8*
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum. Available here
- McCoach DB, Gable RK, Madura JP. (2013). *Instrument development in the affective domain: School and corporate applications*. 3rd. New York: Springer;
- Nápoles, Anna & Santoyo-Olsson, Jasmine & O'Brien, Helen & Stewart, Anita. (2006). *Using Cognitive Interviews to Develop Surveys in Diverse Populations*. *Medical care*. 44. S21-30. [10.1097/01.mlr.0000245425.65905.1d](https://doi.org/10.1097/01.mlr.0000245425.65905.1d).
- Nering, Michael L.; Ostini, Remo, eds. (2010). *Handbook of polytomous item response theory models*. Taylor & Francis. ISBN 978-0-8058-5992-8.
- Ostini, Remo; Nering, Michael L. (2005). *Polytomous Item Response Theory Models. Quantitative Applications in the Social Sciences*. Vol. 144. SAGE. ISBN 978-0-7619-3068-6.
- Polit, D.F. and Beck, C.T. (2006) *Using Research in Evidence-Based Nursing Practice. Essentials of Nursing Research. Methods, Appraisal and Utilization*, Lippincott Williams & Wilkins, Philadelphia, 457-494.
- Rindskopf D., (2001). *International Encyclopedia of the Social & Behavioral Sciences*
- Ruslan Jabrayilov, Wilco H. M. Emons, and Klaas Sijtsma.(2016). *Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment* Published online. doi: [10.1177/0146621616664046](https://doi.org/10.1177/0146621616664046)
- Schmitt, N. (1996). *Uses and abuses of coefficient alpha*. *Psychological Assessment*, 8(4), 350–353. doi:[10.1037/1040-3590.8.4.350](https://doi.org/10.1037/1040-3590.8.4.350).
- Thissen, D., & Wainer, H. (Eds.). (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum. Available here and here
- Willis, GB & Artino, AR (2013). *What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys*. *J Grad Med Educ* 5, 353–356.
- Wu, M., Tam, H. P., & Jen, T-H. (2016). *Educational Measurement for Applied Researchers: Theory into Practice*. Singapore: Springer.

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, 31(4), 231–237. <https://doi.org/10.1027/1015-5759/a000309>

Κασσωτάκης, Μιχάλης Ι., (1999). Η αξιολόγηση της επίδοσης των μαθητών Θεωρητικές προσεγγίσεις και πρακτικές εφαρμογές

Παζινού Ευγενία, Σγουρού Αργυρώ, Παναγιωτακόπουλος Χρήστος, Βερύκιος Βασίλειος. (2017). Η ΘΕΩΡΙΑ ΤΗΣ ΑΠΟΚΡΙΣΗΣ ΕΡΩΤΗΜΑΤΟΣ ΓΙΑ ΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΕΠΙΔΟΣΗΣ ΧΡΗΣΤΩΝ ΕΙΚΟΝΙΚΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΒΙΟΛΟΓΙΑΣ doi: 10.12681/jode.14618

ΠΑΡΑΡΤΗΜΑ Α

ΚΩΔΙΚΑΣ & ΠΑΚΕΤΟ

Στο πλαίσιο της ανάπτυξης κώδικα για την αξιοποίηση, την ερμηνεία και την αξιολόγηση των αποτελεσμάτων, η οποία βασίστηκε στην βιβλιοθήκη

Latent Trait Models under IRT

Version: 1.2-0

Date: 2022-02-18

Author: Dimitris Rizopoulos <d.rizopoulos@erasmusmc.nl>

URL <https://github.com/drizopoulos/lm>

Package Version: 1.2-0 **Date:** 2022-02-18 **License:** GPL

Date/Publication 2022-02-18 09:40:02 UTC

Το πακέτο ltm της R, αναπτύχθηκε για την ανάλυση πολυμεταβλητών διχοτόμων και πολυτομικών δεδομένων με χρήση μοντέλων λανθάνουσας μεταβλητής, σύμφωνα με την προσέγγιση Θεωρίας Απόκρισης Στοιχείων. Για διχοτομικά δεδομένα έχουν εφαρμοστεί το Rasch, το Two-Parameter Logistic και το Birnbaum τριών παραμέτρων, ενώ για πολυτομικά δεδομένα είναι διαθέσιμο το μοντέλο Semejima Graded Response. Οι εκτιμήσεις παραμέτρων λαμβάνονται με οριακή μέγιστη πιθανότητα χρησιμοποιώντας τον κανόνα του τετραγωνισμού Gauss-Ermite.

Τα μοντέλα λανθάνουσας μεταβλητής (Bartholomew et al., 2002) αποτελούν μια γενική κατηγορία μοντέλων κατάλληλων για την ανάλυση πολυμεταβλητών δεδομένων. Κατ' αρχήν, τα μοντέλα λανθάνουσας μεταβλητής είναι μοντέλα πολυμεταβλητής παλινδρόμησης που συνδέουν συνεχείς ή κατηγορικές αποκρίσεις με μη παρατηρούμενες συμμεταβλητές. Οι βα-

σικές παραδοχές και οι στόχοι της μοντελοποίησης λανθάνουσας μεταβλητής μπορεί να συ- νομιστεί ως εξής (Bartholomew et al., 2002)

Ένα μικρό σύνολο λανθάνοντων μεταβλητών θεωρείται ότι εξηγεί τις αλληλεπιδράσεις σε ένα σύνολο παρατηρούμενων μεταβλητών απόκρισης. Αυτό είναι γνωστό ως υπόθεση ανεξαρτησίας υπό όρους, η οποία υποθέτει ότι οι μεταβλητές απόκρισης είναι ανεξάρτητες δεδομένων των λανθάνοντων μεταβλητών. Αυτό απλοποιεί τη διαδικασία εκτίμησης, καθώς η συνεισφορά πιθανότητας των πολυμεταβλητών απαντήσεων αποσυντίθεται σε γινόμενο ανεξάρτητων όρων. Επιπλέον, η διερεύνηση της ανεξαρτησίας υπό όρους μπορεί να βοηθήσει τους ερευνητές πρώτα στην εξαγωγή συμπερασμάτων σύνθετες καταστάσεις και δεύτερον στη σύνοψη των πληροφοριών από τις παρατηρούμενες μεταβλητές σε λίγες διαστάσεις (μείωση διαστάσεων).

- Μη παρατηρούμενες μεταβλητές όπως η ευφυΐα, η μαθηματική ή λεκτική ικανότητα, η φυλετική προκατάληψη, η πολιτική στάση, οι προτιμήσεις των καταναλωτών, που δεν μπορούν να μετρηθούν με συμβατικά μέσα, μπορούν να ποσοτικοποιηθούν υποθέτοντας λανθάνουσες μεταβλητές. Αυτό είναι ένα ελκυστικό χαρακτηριστικό που έχει εφαρμογές σε τομείς όπως το εκπαιδευτικό τεστ, η ψυχολογία, η κοινωνιολογία και το μάρκετινγκ, στους οποίους τέτοιες κατασκευές παίζουν πολύ σημαντικό ρόλο.
- Η μοντελοποίηση λανθάνουσας μεταβλητής χρησιμοποιείται επίσης για την εκχώρηση βαθμολογιών σε δειγματοληπτικές μονάδες στις λανθάνουσες διαστάσεις με βάση τις αποκρίσεις τους. Αυτή η βαθμολογία (γνωστή και ως «Βαθμολογία παράγοντα») είναι μια αριθμητική τιμή που υποδεικνύει τη σχετική απόσταση ή τη θέση ενός ατόμου σε μια λανθάνουσα μεταβλητή. Οι βαθμολογίες παραγόντων μπορούν να χρησιμοποιηθούν είτε για την ταξινόμηση θεμάτων είτε στη θέση του αρχικές μεταβλητές σε μια ανάλυση παλινδρόμησης, με την προϋπόθεση ότι δεν έχει χαθεί η σημαντική διακύμανση στα αρχικά δεδομένα.

Το IRT (Baker, 2004) εξετάζει μια κατηγορία μοντέλων λανθάνουσας μεταβλητής που συνδέουν κυρίως διχοτομικές και πολυτομικές προφανείς μεταβλητές (δηλαδή απόκριση) με μια μεμονωμένη λανθάνουσα μεταβλητή. Οι κύριες εφαρμογές του IRT (Baker et al., 2004) μπορούν να βρεθούν σε εκπαιδευτικές δοκιμασίες στις οποίες οι αναλυτές ενδιαφέρονται να μετρήσουν την ικανότητα των εξεταζόμενων χρησιμοποιώντας ένα τεστ που αποτελείται από πολλά στοιχεία (δηλ. ερωτήσεις). Έχουν προταθεί αρκετά μοντέλα και διαδικασίες εκτίμησης που ασχολούνται με διάφορες πτυχές των εκπαιδευτικών δοκιμασιών.

Ο στόχος αυτής της ενότητας είναι να παρουσιάσει το πακέτο R διαθέσιμο από το CRAN (<http://CRAN.R-project.org/>), το οποίο μπορεί να χρησιμοποιηθεί για να χωρέσει ένα σύνολο μοντέλων λανθάνουσας μεταβλητής στο πλαίσιο της προσέγγισης IRT. Η κύρια εστίαση του πακέτου είναι στα διχοτομικά και πολυτομικά δεδομένα απόκρισης.

Εν συντομία στις κυριότερες εντολές του κώδικα εμπεριέχονται:

Προσαρμογή στο Rasch μοντέλο υπό την προσέγγιση της Θεωρίας Απόκρισης Ερωτήματος

```
rasch(data, constraint = NULL, IRT.param = TRUE, start.val = NULL, na.action = NULL, control = list(), Hessian = TRUE)
```

Προσαρμογή στο Δύο παραμέτρων μοντέλο (2PL) υπό την προσέγγιση της Θεωρίας Απόκρισης Ερωτήματος

```
ltm(formula, constraint = NULL, IRT.param, start.val, na.action = NULL, control = list())
```

Προσαρμογή στο Τριών Παραμέτρων μοντέλο (3PL) υπό την προσέγγιση της Θεωρίας Απόκρισης Ερωτήματος

```
tpm(data, type = c("latent.trait", "rasch"), constraint = NULL, max.guessing = 1, IRT.param = TRUE, start.val = NULL, na.action = NULL, control = list())
```

Εξαγωγή παραμέτρων εκτίμησης

```
coef(object, standardized = FALSE, prob = FALSE, order = FALSE, ...)
```

Καλή Προσαρμογή στο μοντέλο

```
Item.fit()
```

Τεστ Πιθανοφάνειας μεταξύ δύο μοντέλων

```
anova(object, object2, ...)
```

Σχεδιαγράμματα-Γραφικές

```
plot(x, items = NULL, includeFirstLast = FALSE, xlab, ylab, ...)
```

Για τις Χαρακτηριστικές Καμπύλες

plot(x, type = c("ICC", "IIC", "loadings"), items = NULL, zrange = c(-3.8, 3.8), z = seq(zrange[1], zrange[2], length = 100), annot, labels = NULL, legend = FALSE, cx = "top-left", cy = NULL, ncol = 1, bty = "n", col = palette(), lty = 1, pch, xlab, 75 44 plot IRT ylab, zlab, main, sub = NULL, cex = par("cex"), cex.lab = par("cex.lab"), cex.main = par("cex.main"), cex.sub = par("cex.sub"), ask = TRUE, cex.axis = par("cex.axis"), plot = TRUE, ...)

Υπολογισμός Factor Score για το μοντέλο

factor.scores(object, ...)

ΠΑΡΑΡΤΗΜΑ Β

B.1 Ποσοτικά και Κατηγορικά Δεδομένα

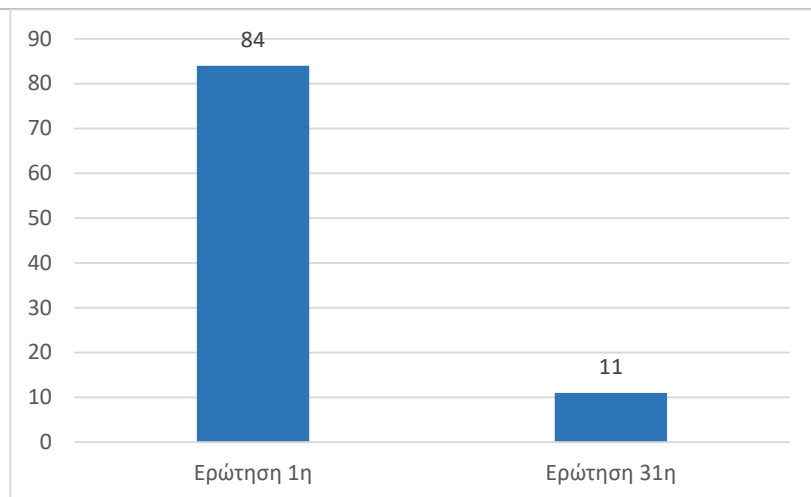
B.2 Σύγκριση ανά ερώτηση

ΣΥΓΚΡΙΤΙΚΟΙ ΠΙΝΑΚΕΣ

B.1 Ποσοτικά & Κατηγορικά Δεδομένα

Στην πρώτη ενότητα του Παραρτήματος Β' παρουσιάζονται ποσοτικά και κατηγορικά δεδομένα που συλλέχθηκαν (ανά κατηγορία-Φοιτητές και Μαθητές) από τα ερωτηματολόγια. Όπως πλήθος, πλήθος ανά φύλο, έτος φοίτησης, η ερώτηση που συγκέντρωσε τις περισσότερες λανθασμένες απαντήσεις, η ερώτηση που συγκέντρωσε τις περισσότερες σωστές απαντήσεις.

Φοιτητές n=98	
Έτος Φοίτησης 1ο (n=97) και 2ο (n=1)	
Θύλη =90 Άρρεν=18	
Μεγαλύτερη Βαθμολογία	28 στα 32
Μικρότερη Βαθμολογία	7 στα 32
Οι ερωτήσεις που συγκέντρωσαν το μεγαλύτερο και το μικρότερο πλήθος σωστών απαντήσεων αντίστοιχα	



1^η ερώτηση | Συντελεστής Δυσκολίας = -0.77528010

13^η ερώτηση | Συντελεστής Δυσκολίας = 5.30554217

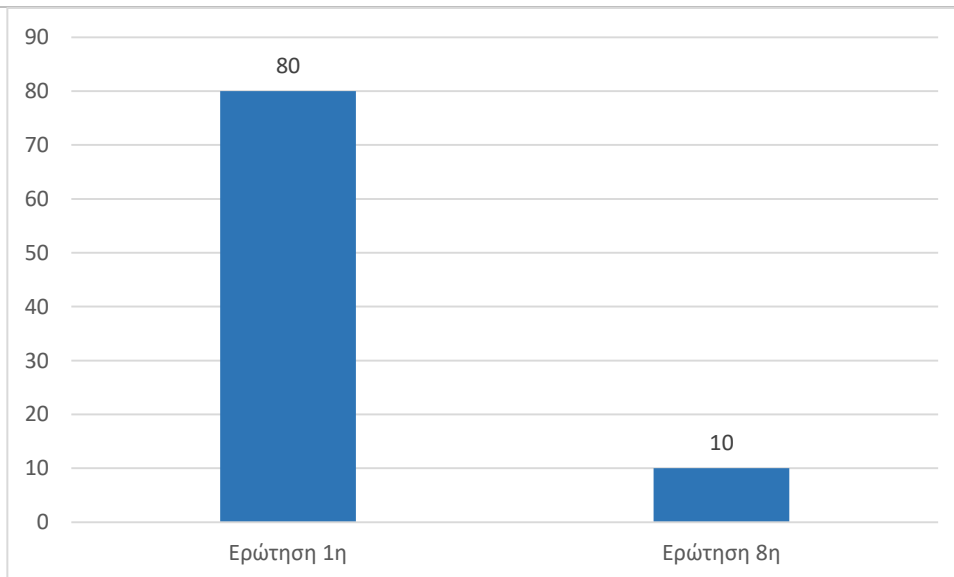
Μαθητές n=123

Θύλη =57 Άρρεν=66

Μεγαλύτερη Βαθμολογία | 21 στα 25

Μικρότερη Βαθμολογία | 2 στα 25

Οι ερωτήσεις που συγκέντρωσαν το μεγαλύτερο και το μικρότερο πλήθος σωστών απαντήσεων αντίστοιχα



1^η ερώτηση | Συντελεστής Δυσκολίας = -1.98933065

8^η ερώτηση | Συντελεστής Δυσκολίας = 1.58122831

B.2 Σύγκριση Ανά ερώτηση

Σε αυτή την ενότητα του Παραρτήματος Β' παρουσιάζονται ραβδογράμματα με στόχο την συγκριτική απεικόνιση του ποσοστού επί τοις εκατό, των σωστών απαντήσεων στις κοινές ερωτήσεις που συνυπάρχουν τόσο στο ερωτηματολόγιο των Φοιτητών όσο και των Μαθητών. Η σειρά των ερωτήσεων είναι- η αύξουσα σειρά των ερωτήσεων όπως δίνονται στο Ερωτηματολόγιο που αφορά τους μαθητές. Στον κάθετο άξονα υπάρχει το ποσοστό των σωστών απαντήσεων με στρογγυλοποίηση στα δέκατα.

Φοιτητές	
Dffclt	
#	Coef
Q1	-1.98933065
Q2	-1.13693031
Q3	-0.56287531
Q4	0.10084029
Q5	0.84362568
Q6	-1.50630494
Q7	19.54868471
Q8	1.58122831
Q9	0.65698346
Q10	-0.30838180
Q11	0.50365888
Q12	-0.71633344
Q13	-0.94358987
Q14	-0.35711345
Q15	-1.63574439
Q16	1.09535169
Q17	-1.49869332
Q18	2.26774353
Q19	1.51860105
Q20	2.85758781
Q21	-0.87076424
Q22	-0.05870479
Q23	-0.23778640
Q24	-0.18745453
Q25	-6.80514445

Μαθητές	
Dffclt	
#	Coef
Q1	-0.77528010
Q2	-2.40787048
Q3	1.38301960
Q4	1.46063586
Q5	2.33421667
Q6	0.14773760
Q7	-1.93590247
Q8	-5.83620268
Q9	36.50302952
Q10	3.15857862
Q11	1.60656085
Q12	0.81038074
Q13	5.30554217
Q14	-0.18663998
Q15	-21.79700233
Q16	7.91930995
Q17	-0.72158988
Q18	7.07276331
Q19	-7.11171806
Q20	0.89618260
Q21	-0.06414096
Q22	-6.62668200
Q23	1.08749328
Q24	2.06972814
Q25	5.88848037

#	Ποσοστό μαθητών επι % που απάντησαν σωστά	Ποσοστό φοιτητών επι % που απάντησαν σωστά
7	25	14
15	24	13
18	27	17
20	36	31
25	32	16

Στον τρίτο πίνακα παρουσιάζονται οι τέσσερις ερωτήσεις στις οποίες οι μαθητές είχαν ξεκάθαρο προβάδισμα. Μελετώντας τον πρώτο πίνακα οι ερωτήσεις Q7 και Q20 έχουν ξεκάθαρα μεγάλο συντελεστή δυσκολίας για τους φοιτητές ενώ για τους μαθητές η ερώτηση Q7 και Q20 δεν παρουσιάζουν τόσο μεγάλη δυσκολία. Η Q15 παρότι δεν έχει μεγάλη δυσκολία για τους φοιτητές έχει πολύ μικρή δυσκολία για τους μαθητές. Η ερώτηση 18 ήταν δύσκολη για τους φοιτητές αλλά σαφώς ακόμη δυσκολότερη για τους μαθητές αλλά παρόλα αυτά απάντησαν καλύτερα.

Τέλος, η Q25 έχει αρνητικό πρόσημο οπότε έπρεπε θεωρητικά να είναι εύκολη για τους φοιτητές ενώ για τους μαθητές έχει πολύ μεγάλο συντελεστή δυσκολίας πάλι οι μαθητές έχουν προβάδισμα. Αυτό, που μπορεί να συμπεράνει κάποιος-αν θέλει να κάνει μια άτυπη σύγκριση-μιας και πολλοί παράγοντες παίζουν ρόλο όπως το μικρό μέγεθος των δύο δειγμάτων, το αντικείμενο των ερωτήσεων είχε δομηθεί με βάση το περιεχόμενο του βιβλίου των Μαθηματικών της Στ' Δημοτικού-επομένως μπορούν να θεωρηθούν πιο «πρόσφατες» γνώσεις για τους μαθητές της Α' γυμνασίου άρα και να απαντούν με μεγαλύτερη ευκολία από τους φοιτητές και σαφώς ο παράγοντας τύχης που πάντα εισβάλλει σε έρευνες.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Ρίζου Αρσινόη Ειρήνη γεννήθηκε στα Ιωάννινα. Σπούδασε στο τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων, συνέχισε με μεταπτυχιακές σπουδές στο τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής με ειδίκευση στους Επιστημονικούς Υπολογισμούς και είναι τελειόφοιτη φοιτήτρια του Π.Μ.Σ. του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης του Πανεπιστημίου Ιωαννίνων. Έχει εργαστεί ως Μαθηματικός σε ιδιωτικές δομές εκπαίδευσης και δημόσια σχολεία αλλά και ως Συνεργάτης Έρευνας στο Εθνικό Κέντρο Έρευνας και Τεχνολογικής Ανάπτυξης. Τα ερευνητικά της ενδιαφέροντα εστιάζουν στην Ανάλυση και Μετανάλυση δεδομένων, την διδακτική των Μαθηματικών πρωτοβάθμιας και δευτεροβάθμιας Εκπαίδευσης και τις Νέες Τεχνολογίες στην Εκπαίδευση.