

## ΕΞΕΤΑΣΗ ΤΩΝ ΨΥΧΟΜΕΤΡΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΡΓΑΛΕΙΩΝ ΜΕΤΡΗΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΑΠΟΚΡΙΣΗΣ ΣΤΟΙΧΕΙΟΥ: ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΣΤΗΝ ΠΡΑΞΗ

*Νικόλαος Τσιγγίλης*

*Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης*

**Περίληψη:** Η Θεωρία Απόκρισης Στοιχείου (ΘΑΣ) αποτελεί μιαν εναλλακτική προσέγγιση της Κλασικής Θεωρίας Μέτρησης (ΚΘΜ) για την ανάπτυξη εργαλείων μέτρησης και την εξέταση των ψυχομετρικών τους ιδιοτήτων. Η ΘΑΣ, όπως προκύπτει από το όνομά της, εστιάζεται και εξετάζει το κάθε στοιχείο του εργαλείου μέτρησης. Σκοπός της είναι να μοντελοποιήσει τη σχέση μεταξύ του λανθάνοντος χαρακτηριστικού των συμμετεχόντων (latent trait) και της απόκρισής των στα διάφορα στοιχεία του εργαλείου μέτρησης. Στην παρούσα εργασία παρουσιάζονται οι βασικές έννοιες της ΘΑΣ, τα πλεονεκτήματά της συγκριτικά με την ΚΘΜ, οι προϋποθέσεις εφαρμογής της καθώς ο τρόπος ερμηνείας και αξιοποίησης των αποτελεσμάτων σε μια προσπάθεια ευαισθητοποίησης των ερευνητών για τη χρησιμότητά της και τις εφαρμογές της. Η παρουσίαση εστιάζεται στα πιο απλά μοντέλα όπου χρησιμοποιούνται κυρίως διχοτομικά δεδομένα καθώς η σε βάθος κατανόησή των διευκολύνει την ομαλή μετάβαση στην εφαρμογή πιο σύνθετων τεχνικών. Στη συνέχεια παρουσιάζεται μια εφαρμογή της ανάλυσης με τη χρήση δύο στατιστικών προγραμμάτων, ενός εμπορικού με γραφικό περιβάλλον (IRTPRO) και ενός ελεύθερου λογισμικού (Irm) στο περιβάλλον της R.

**Λέξεις κλειδιά:** Θεωρία απόκρισης στοιχείου, Μοντέλα λανθανουσών μεταβλητών, Ψυχομετρικές ιδιότητες δοκιμασιών

---

**Διεύθυνση:** Νικόλαος Τσιγγίλης, Τμήμα Δημοσιογραφίας και ΜΜΕ, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Εγνατία 46, 54625 Θεσσαλονίκη. Tel: 2310-991997. E-mail: ntsigilis@jour.auth.gr

## ΕΙΣΑΓΩΓΗ

Αρκετές από τις κοινωνικές επιστήμες (π.χ., παιδαγωγική, ψυχολογία, κοινωνιολογία) συχνά ασχολούνται με εννοιολογικές κατασκευές, όπως, για παράδειγμα, οι στάσεις των ατόμων ως προς την ανακύκλωση, τα κίνητρα των μουσικών προτιμήσεων, η κατάθλιψη, η επαγγελματική εξουθένωση εργαζόμενων κ.ά. Στην περίπτωση αυτή η μέτρησή τους δε γίνεται άμεσα, αλλά έμμεσα μέσω των εκφάνσεών τους. Πρόκειται, δηλαδή, για λανθάνουσες μεταβλητές. Μια πρόκληση των ερευνητών που δραστηριοποιούνται στο χώρο των κοινωνικών επιστημών είναι να τεκμηριώσουν ότι πράγματι μπορούν να μετρήσουν αυτές τις εννοιολογικές κατασκευές. Χωρίς την παροχή ισχυρών εμπειρικών ευρημάτων τα οποία να συνηγορούν ως προς τη δυνατότητα κατανόησης και αποτύπωσης των εννοιολογικών αυτών κατασκευών δυσχεραίνεται η μελέτη και η εξέλιξη θεωριών και μειώνεται η εμπιστοσύνη της αποτελεσματικότητας τυχόν παρεμβάσεων.

Για αρκετά χρόνια το κυρίαρχο θεωρητικό πλαίσιο για την ανάπτυξη και αξιολόγηση των ψυχομετρικών χαρακτηριστικών εργαλείων με σκοπό την αποτύπωση των εννοιολογικών κατασκευών ήταν η Κλασική Θεωρία Μέτρησης (ΚΘΜ, Classical Test Theory) (π.χ., Embretson & Reise, 2000; Edwards, 2009). Βασική αρχή της ΚΘΜ είναι ότι η κάθε παρατηρούμενη τιμή είναι το άθροισμα της αληθινής τιμής του χαρακτηριστικού του ατόμου και του σφάλματος της μέτρησης (π.χ., Hambleton & Jones, 1993. Mango, 2009). Έτσι, κύρια επιδίωξη είναι η μείωση του σφάλματος μέτρησης. Αναλύσεις όπως διερευνητική παραγοντική ανάλυση, αξιοπιστία εσωτερικής συνοχής, συσχέτιση στοιχείου-συνολικής βαθμολογίας στηρίζονται στην ΚΘΜ.

Τα τελευταία χρόνια όμως το ενδιαφέρον των επιστημόνων αρχίζει να στρέφεται στη Θεωρία Απόκρισης Στοιχείου (Item Response Theory) (Edwards, 2009. Toland, 2014). Η στροφή αυτή στηρίζεται στα ισχυρά πλεονεκτήματά της έναντι της ΚΘΜ. Όπως φαίνεται και από το όνομά της, η ΘΑΣ μελετάει το κάθε ένα στοιχείο του εργαλείου μέτρησης<sup>1</sup>. Το επίκεντρο της συγκεκριμένης προσέγγισης είναι στη σχέση μεταξύ των παραμέτρων του κάθε στοιχείου του εργαλείου μέτρησης και της λανθάνουσας έννοιας (theta,  $\theta$ ) (DeMars, 2010).

Σκοπός του άρθρου είναι να εισαγάγει τους ερευνητές που δραστηριοποιούνται στο χώρο των κοινωνικών επιστημών στη Θεωρία Απόκρισης Στοιχείου. Πιο συγκεκριμένα, επιχειρεί να αναδείξει τα πλεονεκτήματα της ΘΑΣ έναντι της ΚΘΜ, τις

<sup>1</sup> Με τη λέξη «στοιχείο» αποδίδεται στα ελληνικά η αγγλική λέξη item. Επιλέχθηκε η συγκεκριμένη απόδοση με προτροπή του Καθηγητή Ιωάννη Τσαούση, ως πιο γενική και αντιπροσωπευτική για τις διάφορες περιπτώσεις που μπορεί να χρησιμοποιηθεί. Έτσι, στοιχείο μπορεί να είναι: α) η πρόταση ενός ερωτηματολογίου, β) η κινητική δοκιμασία μιας δέσμης κινητικών μετρήσεων, γ) η ερώτηση πολλαπλής επιλογής σε ένα τεστ γνώσεων κ.ά.

προϋποθέσεις εφαρμογής της, και να παρουσιάσει ένα ολοκληρωμένο παράδειγμα με τη χρήση δύο στατιστικών προγραμμάτων, ενός εμπορικού (IRTPRO ver. 4.2, Cai, du Toit, & Thissen, 2017) και ενός ανοικτού κώδικα (library ltm της R, Rizopoulos, 2006). Η παρουσίαση καθώς και η πρακτική εφαρμογή θα εστιαστεί στα πιο απλά μοντέλα όπου χρησιμοποιούνται κυρίως διχοτομικά δεδομένα. Η σε βάθος κατανόηση των διχοτομικών μοντέλων είναι αρκετά σημαντική καθώς διευκολύνει την ομαλή και αποτελεσματική μετάβαση στην εφαρμογή πιο σύνθετων τεχνικών, αφού οι βασικές έννοιες δε διαφοροποιούνται.

### **Σύγκριση της Κλασικής Θεωρίας Μέτρησης και της Θεωρίας Απόκρισης Στοιχείου**

Στο τμήμα αυτό θα αναφερθούν κάποιες από τις διαφορές μεταξύ των δύο θεωρητικών πλαισίων έτσι ώστε να γίνει αντιληπτή η χρησιμότητα της ΘΑΣ. Κατά τη στιγμή της χορήγησης του εργαλείου μέτρησης (π.χ., τεστ μαθηματικών γνώσεων, ερωτηματολόγιο επαγγελματικής εξουθένωσης) ο συμμετέχων διαθέτει ένα συγκεκριμένο επίπεδο λανθάνοντος χαρακτηριστικού ή ικανότητας. Εάν το εργαλείο μέτρησης περιέχει δύσκολα στοιχεία τότε η επίδοση αναμένεται να είναι χαμηλή. Στην περίπτωση όμως που τα στοιχεία είναι εύκολα η επίδοση θα είναι υψηλή. Συνεπώς, η έννοια του λανθάνοντος χαρακτηριστικού είναι πιο γενική και ανεξάρτητη από τη χορηγούμενη δοκιμασία, ενώ η αληθινή τιμή είναι πιο ειδική και διαφοροποιείται ανάλογα με τη δοκιμασία (Hambleton & Jones, 1993). Το πλεονέκτημα της ΘΑΣ είναι ότι μπορεί να διαχωρίσει το επίπεδο λανθάνοντος χαρακτηριστικού από τις παραμέτρους του στοιχείου (π.χ., δυσκολία) παρέχοντας μία εκτίμηση η οποία δεν επηρεάζεται από το συγκεκριμένο δείγμα που χρησιμοποιήθηκε. Η ανεξαρτησία των παραμέτρων των στοιχείων (και, κατά συνέπεια, και του εργαλείου μέτρησης) από τα χαρακτηριστικά του δείγματος που χρησιμοποιήθηκε είναι βασικό ζητούμενο στο χώρο της μέτρησης και της αξιολόγησης.

Μία άλλη σημαντική διαφοροποίηση μεταξύ της ΚΘΜ και της ΘΑΣ είναι η αντίληψη για το σφάλμα μέτρησης και η επίδρασή του στην αξιολόγηση της αξιοπιστίας του εργαλείου μέτρησης. Στην ΚΘΜ το σφάλμα μέτρησης και, κατά συνέπεια, η αξιοπιστία, θεωρείται ότι είναι σταθερά και ανεξάρτητα από το επίπεδο ικανότητας ή λανθάνοντος χαρακτηριστικού του ατόμου. Συνεπώς, παρέχεται μία και ίδια τιμή για όλο το εργαλείο μέτρησης (π.χ., Cronbach's  $\alpha$ ). Αντίθετα, στη ΘΑΣ το σφάλμα μέτρησης δεν είναι σταθερό αλλά μεταβάλλεται ανάλογα με το επίπεδο του λανθάνοντος χαρακτηριστικού. Η διαφοροποίηση αυτή μπορεί να αποτυπωθεί τόσο για το κάθε στοιχείο όσο και συνολικά για το εργαλείο μέτρησης (π.χ., DeMars, 2010. Magno, 2009) κι επιτρέπει βαθύτερη κατανόηση και πληρέστερη αποτίμηση της αξιοπιστίας.

Σε αντίθεση με τη ΘΑΣ, η ΚΟΜ εστιάζεται στο επίπεδο του εργαλείου μέτρησης (de Ayala, 2009). Η συνολική επίδοση στο εργαλείο μέτρησης αντιπροσωπεύει το λανθάνον χαρακτηριστικό. Η συνολική επίδοση προκύπτει συνήθως από το άθροισμα της βαθμολογίας στα επιμέρους στοιχεία. Η άθροιση όμως αυτή έχει τα εξής μειονεκτήματα: α) είναι παραπλανητική, με την έννοια ότι στην περίπτωση μετρήσεων με κλίμακες Likert οι αποστάσεις δεν μπορεί να θεωρηθούν ως ίσες, και συνεπώς το άθροισμα δεν έχει ουσιαστικό νόημα (Zhu, 1996)· β) η απόδοση συγκεκριμένων αριθμών σε χαρακτηρισμούς (π.χ., συμφωνώ απόλυτα) είναι αυθαίρετη. Για παράδειγμα, αντί για κλίμακα 1 - 5 θα μπορούσε να χρησιμοποιηθεί κλίμακα 0 - 4 (Torpe & Favia, 2012), και γ) προϋποθέτει ότι η βαρύτητα όλων των στοιχείων στη διαμόρφωση του λανθάνοντος χαρακτηριστικού είναι ίδια. Για παράδειγμα, το Self-Harm Inventory (SHI, Sansone, Wiederman, & Sansone, 1998) περιέχει 22 στοιχεία για τη διερεύνηση του ιστορικού αυτοτραυματισμού, με τις απαντήσεις να δίνονται σε δίτιμη κλίμακα, Ναι - Όχι. Σύμφωνα με τους κατασκευαστές, η συνολική τιμή προκύπτει από το άθροισμα των ερωτήσεων που απαντήθηκαν θετικά, και τιμές από πέντε και πάνω είναι ένδειξη σοβαρού προβλήματος. Τρία από τα στοιχεία του SHI είναι τα εξής: «Έχετε χαρακώσει τον εαυτό σας σκόπιμα;» (#2), «Έχετε οδηγήσει ηθελημένα απερίσκεπτα;» (#7), και «Έχετε επιχειρήσει αυτοκτονία;» (#18). Είναι φανερό ότι θετική απάντηση στο στοιχείο #18 θα πρέπει να έχει μεγαλύτερη βαρύτητα από ότι στα στοιχεία #2 και #7. Αντιθέτως, στη ΘΑΣ κατά τον υπολογισμό της λανθάνουσας ικανότητας δε λαμβάνεται μόνο υπόψη πόσα στοιχεία απαντήθηκαν θετικά αλλά και ποια ήταν αυτά. Είναι, δηλαδή, δυνατόν δύο συμμετέχοντες που απάντησαν θετικά σε πέντε στοιχεία του SHI να έχουν διαφορετικές τιμές λανθάνοντος χαρακτηριστικού, επειδή τα στοιχεία στα οποία απάντησαν θετικά είχαν διαφορετικά χαρακτηριστικά (π.χ., διαφορετικές τιμές διάκρισης και δυσκολίας). Με τον τρόπο αυτό δημιουργείται μία συνολική τιμή πιο ευαίσθητη και πιο αξιόπιστη από ότι το άθροισμα των ορθών απαντήσεων (DeMars, 2010) (Βλ. σελίδες 12-13 Παραρτήματος Β). Η αποτύπωση του λανθάνοντος χαρακτηριστικού υπό το πρίσμα της ΘΑΣ έχει ως συνέπεια τη λεπτομερέστερη κατάταξη και αξιολόγηση των συμμετεχόντων καθώς και την ορθότερη διατύπωση συμπερασμάτων κατά την εξέταση άλλων ψυχομετρικών ιδιοτήτων όπως, π.χ., η εγκυρότητα κριτηρίου.

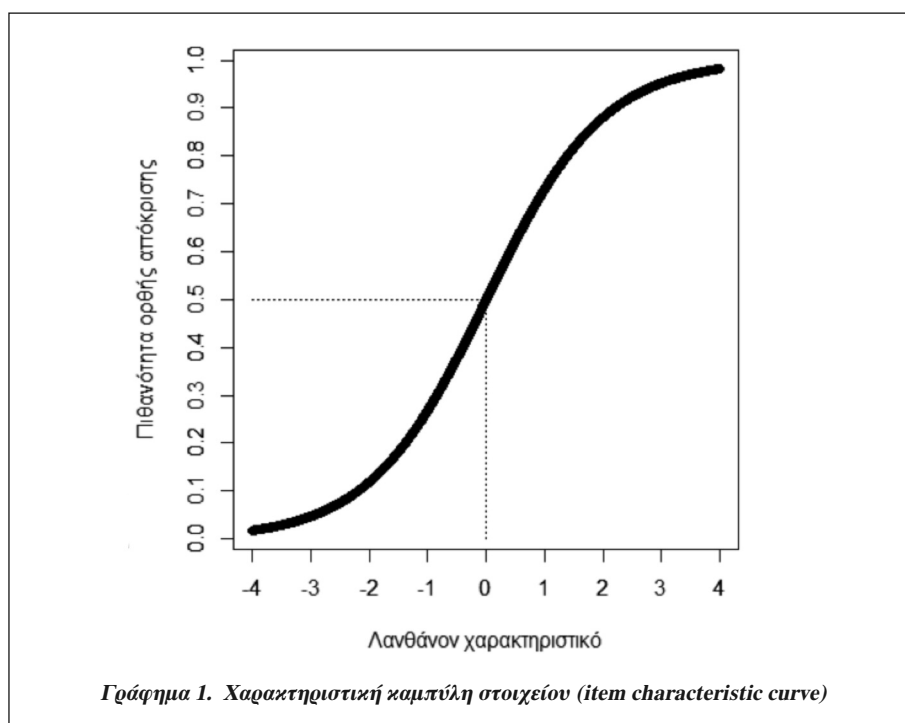
### **Θεωρία Απόκρισης Στοιχείου – λογαριθμικά μοντέλα**

Αντίθετα από ό,τι υπονοεί το όνομά της, η ΘΑΣ είναι περισσότερο μία ομάδα μαθηματικών μοντέλων τα οποία συνδέουν τις παρατηρούμενες με τις λανθάνουσες μεταβλητές με ένα μη γραμμικό τρόπο. Η βασική ιδέα της ΘΑΣ είναι ότι η απόκριση σε ένα στοιχείο μιας δέσμης μετρήσεων (π.χ., ερωτηματολόγιο, τεστ γνώσεων) είναι συνάρτηση των χαρακτηριστικών του ατόμου (π.χ., πόσο ισχυρά κίνητρα έχει) και των

ιδιοτήτων του στοιχείου (π.χ., πόσο δύσκολο ή εύκολο είναι). Η χρήση γραφημάτων για την οπτική παρουσίαση των διάφορων μοντέλων της ΘΑΣ έχει μακριά παράδοση. Στο Γράφημα 1 φαίνεται η χαρακτηριστική καμπύλη στοιχείου (item characteristic curve, ICC), που περιγράφει τη σχέση μεταξύ ενός στοιχείου και του λανθάνοντος χαρακτηριστικού. Η καμπύλη αυτή βασίζεται στην παρακάτω συνάρτηση:

$$P(x_j=1|\theta, b_j) = \frac{1}{1 + e^{-(\theta - b_j)}}$$

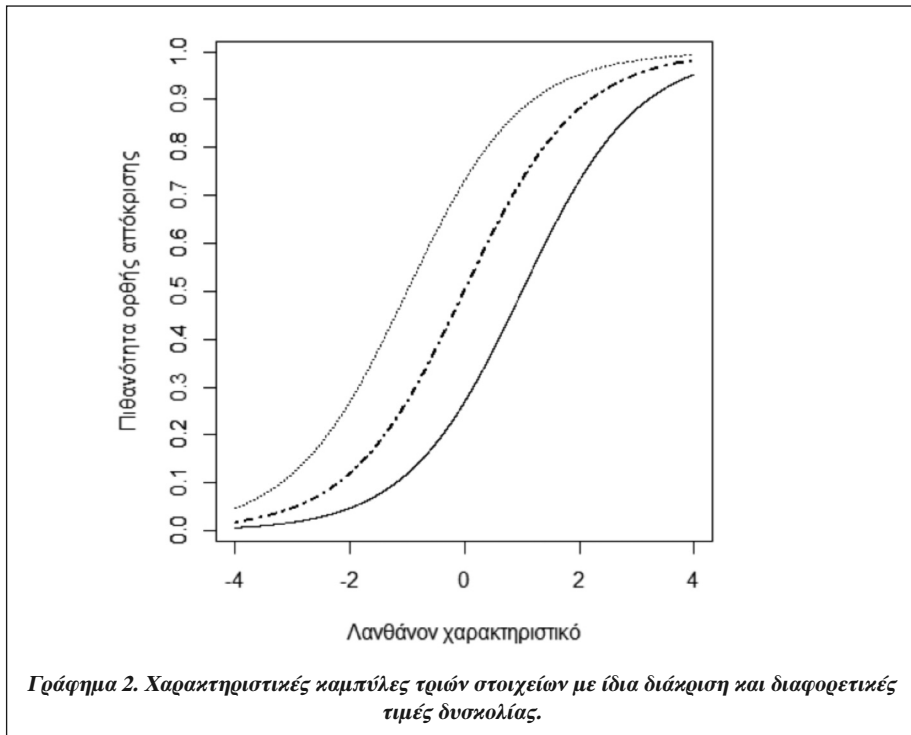
Η ποσότητα  $P(x_j = 1|\theta, b_j)$  δηλώνει την πιθανότητα ορθής απόκρισης<sup>2</sup> στο στοιχείο  $x_j$ , δεδομένου ενός συγκεκριμένου επιπέδου λανθάνοντος χαρακτηριστικού ( $\theta$ , theta). Το  $e$  είναι η βάση του φυσικού λογαριθμού ( $e = 2.71828\dots$ ) και το  $b_j$  είναι η παράμετρος δυσκολίας του στοιχείου (difficulty or location parameter). Ο δείκτης  $j$  της παραμέτρου  $b_j$  δηλώνει ότι η παράμετρος της δυσκολίας δεν είναι ίδια για όλα τα στοιχεία αλλά μπορεί να διαφέρει.



<sup>2</sup> Η έκφραση «πιθανότητα ορθής απόκρισης» αναφέρεται περισσότερο σε τεστ γνώσεων με τη χρήση ερωτήσεων πολλαπλών επιλογών. Στο χώρο άλλων επιστημονικών πεδίων όπως, π.χ., η ψυχολογία μπορεί να εννοηθεί ως «πιθανότητα θετικής απόκρισης» στη συγκεκριμένη πρόταση-δήλωση.

Ο οριζόντιος άξονας αναπαριστά τη λανθάνουσα ικανότητα ή το λανθάνον χαρακτηριστικό, το οποίο συνήθως θεωρείται ότι ακολουθεί την τυπική κανονική κατανομή (Edwards, 2009). Συνεπώς, το μεγαλύτερο ποσοστό των τιμών  $\theta$  αναμένεται να κυμαίνεται από  $-3$  έως  $+3$ . Το σημείο 0 είναι το μέσο επίπεδο του χαρακτηριστικού (ή ικανότητας), ενώ το σημείο  $+1$  δείχνει επίπεδο χαρακτηριστικού που αντιστοιχεί σε μία τυπική απόκλιση πάνω από τη μέση ικανότητα. Στον κατακόρυφο άξονα αναπαρίστανται η πιθανότητα ορθής ή θετικής απόκρισης στο στοιχείο. Οι τιμές κυμαίνονται από 0 έως 1. Από τη χαρακτηριστική καμπύλη στοιχείου φαίνεται ότι σε χαμηλά επίπεδα ικανότητας αντιστοιχεί χαμηλή πιθανότητα επιτυχίας. Όσο αυξάνεται η ικανότητα τόσο αυξάνεται και η πιθανότητα επιτυχίας. Η σχέση αυτή όμως δεν είναι γραμμική. Έτσι, μετά από ένα ορισμένο σημείο αύξησης (ή μείωσης, αντιστοίχως) της ικανότητας προκαλείται μικρή αύξηση (μείωση) της αναμενόμενης επιτυχίας.

Στην περίπτωση ενός διχοτομικού στοιχείου, η παράμετρος δυσκολίας δηλώνει το μέγεθος του λανθάνοντος χαρακτηριστικού που πρέπει να έχει ο συμμετέχων για να υπάρχει 50% πιθανότητα ορθής απόκρισης στο στοιχείο. Όσο πιο μετατοπισμένη δεξιά είναι η χαρακτηριστική καμπύλη ενός στοιχείου τόσο μεγαλύτερο το μέγεθος του λανθάνοντος χαρακτηριστικού που απαιτείται για να υπάρχει 50% πιθανό-



τητα ορθής απόκρισης. Συνεπώς, το στοιχείο είναι πιο δύσκολο. Στο Γράφημα 2 παρουσιάζονται οι χαρακτηριστικές καμπύλες τριών διαφορετικών στοιχείων. Το στοιχείο με διακεκομμένη γραμμή έχει παράμετρο δυσκολίας  $-1$  και βρίσκεται αριστερά των άλλων δύο, και συνεπώς είναι το ευκολότερο. Το στοιχείο με τη συνεχόμενη χαρακτηριστική καμπύλη έχει παράμετρο δυσκολίας  $+1$ , βρίσκεται δεξιότερα των άλλων και συνεπώς είναι αυτό που απαιτεί υψηλότερα επίπεδα λανθάνοντος χαρακτηριστικού για να υπάρχει πιθανότητα 50% ορθής απόκρισης.

Το μοντέλο που περιγράφηκε παραπάνω αναφέρεται ως λογαριθμικό μοντέλο μιας παραμέτρου (1PL logistic model)<sup>3</sup>. Η παράμετρος που διαφοροποιείται ανάλογα με το κάθε στοιχείο είναι η δυσκολία. Ένα άλλο ευρέως χρησιμοποιούμενο μοντέλο στην περίπτωση διχοτομικών δεδομένων είναι το λογαριθμικό μοντέλο δύο παραμέτρων (2PL logistic model)<sup>4</sup> και παριστάνεται από την συνάρτηση:

$$P(x_j=1|\theta, \alpha_j, b_j) = \frac{1}{1 + e^{-\alpha_j(\theta - b_j)}}$$

Η διαφορά από το προηγούμενο μοντέλο έγκειται στην προσθήκη της παραμέτρου διάκρισης (discrimination parameter,  $\alpha_j$ ) στον εκθέτη της βάσης του φυσικού λογάριθμου ( $e$ ). Αξίζει να σημειωθεί ότι ο δείκτης  $j$  δηλώνει ότι η διάκριση δεν είναι ίδια για όλα τα στοιχεία αλλά μπορεί να διαφέρει. Η έννοια της διάκρισης αφορά την κλίση της καμπύλης μεταξύ του λανθάνοντος χαρακτηριστικού και της πιθανότητας ορθής απόκρισης, και συνδέεται στενά με την έννοια της φόρτισης στη διερευνητική παραγοντική ανάλυση (Edwards, 2009). Όσο μεγαλύτερη είναι η κλίση του στοιχείου τόσο υψηλότερη είναι και η τιμή της διάκρισης και τόσο εντονότερη η σχέση με το λανθάνον χαρακτηριστικό. Στοιχεία με υψηλές τιμές διάκρισης μπορούν να διαχωρίσουν με πιο αποτελεσματικό τρόπο άτομα με παραπλήσια επίπεδα λανθάνοντος χαρακτηριστικού. Με πιο απλά λόγια, για ένα στοιχείο με υψηλή τιμή διάκρισης ακόμη και μικρή μεταβολή στο λανθάνον χαρακτηριστικό οδηγεί σε μεγάλες μεταβολές της πιθανότητας ορθής απόκρισης. Οι Morizot, Ainsworth, και Reise (2007) αναφέρουν ότι στο χώρο της μελέτης της προσωπικότητας οι τι-

<sup>3</sup> Αξίζει να διευκρινιστεί ότι στην περίπτωση που η τιμή της διάκρισης οριστεί στη μονάδα ( $\alpha = 1$ ) για όλα τα στοιχεία, το μοντέλο αναφέρεται ως Rasch μοντέλο, ενώ στη περίπτωση που η τιμή της διάκρισης έχει μια συγκεκριμένη τιμή για όλα τα στοιχεία ( $\alpha$ ) αλλά όχι απαραίτητα μονάδα, το μοντέλο αναφέρεται ως μοντέλο μιας παραμέτρου (1PL model).

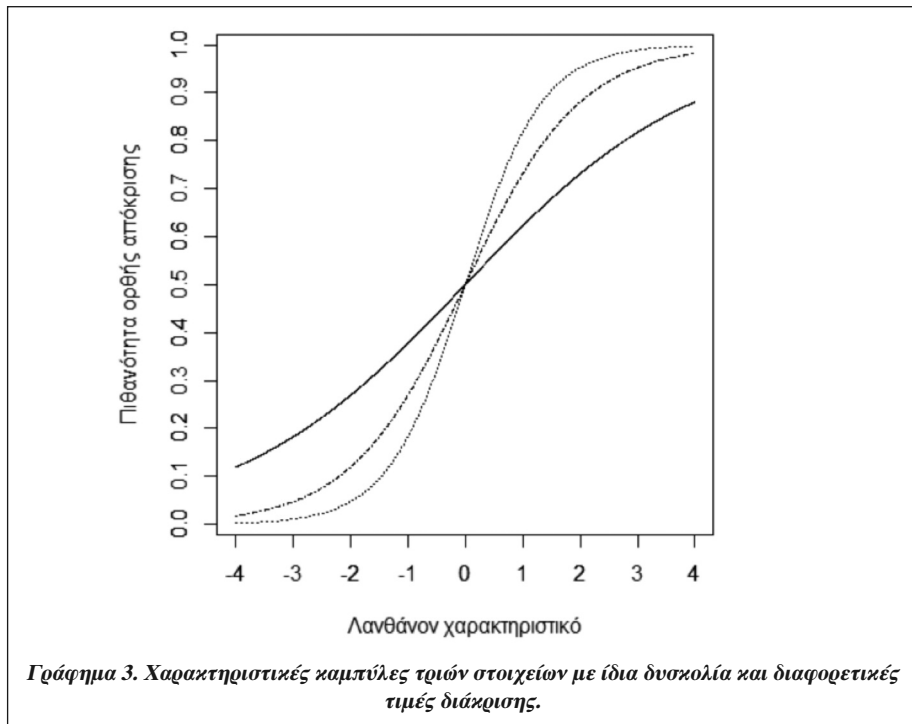
<sup>4</sup> Στη βιβλιογραφία επίσης αναφέρεται και το μοντέλο τριών παραμέτρων (3PL logistic model), όπου εκτός από τη διάκριση και τη δυσκολία υπάρχει και η παράμετρος τυχαίας επιλογής (guessing)  $c_j$ . Η παράμετρος αυτή δηλώνει την πιθανότητα ορθής απάντησης από άτομα χαμηλής ικανότητας (DeMars, 2010). Το μοντέλο αυτό δε θα εξεταστεί για τον περιορισμό της έκτασης της εργασίας και επειδή έχει περισσότερο νόημα σε τεστ γνώσεων.

μές διάκρισης κυμαίνονται από 0.5 έως 2.0, με συνηθέστερες τιμές γύρω από τη μονάδα.

Στο Γράφημα 3 παρουσιάζονται οι χαρακτηριστικές καμπύλες τριών στοιχείων. Για όλα τα στοιχεία η παράμετρος δυσκολίας έχει οριστεί στο μηδέν, και αυτό που τα διαφοροποιεί είναι η παράμετρος διάκρισης. Πιο συγκεκριμένα, το στοιχείο με συνεχή γραμμή έχει  $\alpha = 0.5$ , το στοιχείο με διακεκομμένη γραμμή έχει  $\alpha = 1.5$  και το ενδιάμεσο στοιχείο έχει  $\alpha = 1$ . Ένα παράδειγμα ίσως θα βοηθούσε να γίνει καλύτερα κατανοητή η σημασία της έννοιας της διάκρισης. Έστω δύο συμμετέχοντες με τιμές λανθάνοντος χαρακτηριστικού  $\theta_1 = -.5$  και  $\theta_2 = .5$ . Η πιθανότητα ορθής απόκρισης στο πρώτο στοιχείο ( $\alpha = 0.5$ ) είναι .438 και .562 αντιστοίχως, δηλαδή μια διαφορά .124. Τα ποσοστά ορθής απόκρισης των δύο συμμετεχόντων ως προς το στοιχείο με τιμή διάκρισης 1.5 είναι .321 και .679, δηλαδή μια διαφορά .349, σχεδόν τριπλάσια της πρώτης.

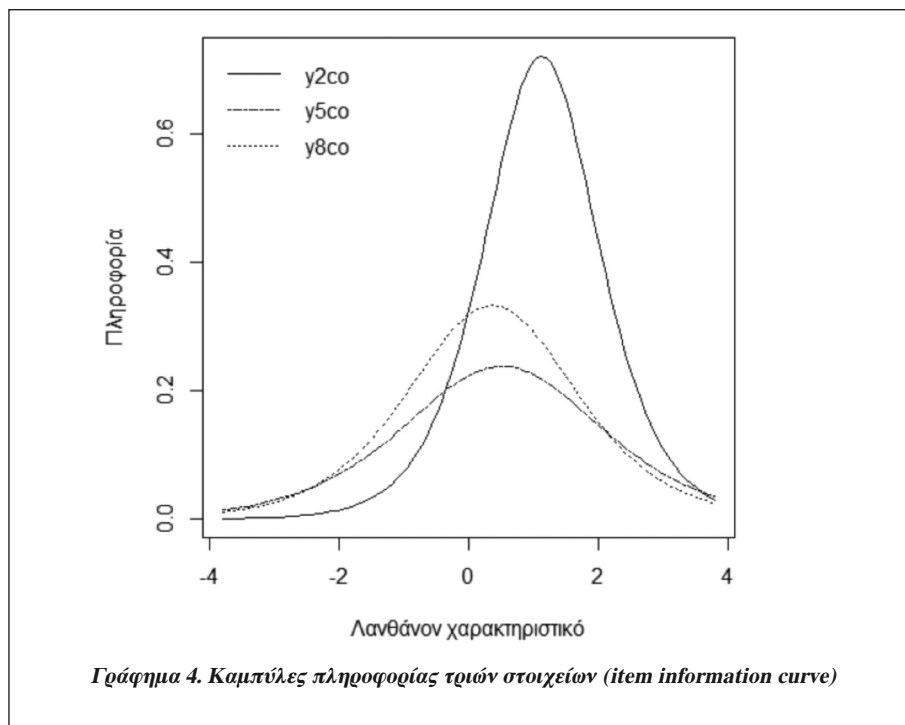
### Πληροφορία στοιχείου και εργαλείου μέτρησης

Μια άλλη σημαντική διαφοροποίηση μεταξύ της ΚΘΜ και της ΘΑΣ έγκειται στην προσέγγιση της έννοιας της αξιοπιστίας. Σύμφωνα με την ΚΘΜ, η αξιοπιστία ανα-





φέρεται στο βαθμό στον οποίο οι παρατηρούμενες τιμές προσεγγίζουν τις πραγματικές τιμές. Η εκτίμηση της αξιοπιστίας γίνεται με τον υπολογισμό μιας τιμής (π.χ., Cronbach's  $\alpha$ , Intraclass Correlation Coefficient), η οποία ισχύει για το σύνολο του εργαλείου μέτρησης, ανεξάρτητα από το επίπεδο του λανθάνοντος χαρακτηριστικού. Αντιθέτως, στη ΘΑΣ υπάρχει η έννοια της «πληροφορίας». Η πληροφορία ορίζεται ως συνάρτηση του τυπικού σφάλματος ( $I = 1/SE^2$ ) (Baker, 2001), δεν είναι καθολική αλλά μεταβάλλεται ανάλογα με τις παραμέτρους του κάθε στοιχείου και ανάλογα με το επίπεδο του λανθάνοντος χαρακτηριστικού. Πιο συγκεκριμένα, το μέγεθος της πληροφορίας καθορίζεται από το βαθμό διάκρισης του στοιχείου, ενώ η περιοχή του λανθάνοντος χαρακτηριστικού στην οποία συγκεντρώνεται η πληροφορία καθορίζεται από τη δυσκολία του στοιχείου. Στο Γράφημα 4 παρουσιάζεται η καμπύλη πληροφορίας στοιχείου (item information curve, IIC) για τρία στοιχεία. Από το γράφημα γίνεται εύκολα κατανοητό ότι και τα τρία στοιχεία αξιολογούν το λανθάνον χαρακτηριστικό με μεγαλύτερη ακρίβεια στην περιοχή από 0 έως 2. Επίσης, το στοιχείο της συνεχούς γραμμής παρέχει τη μεγαλύτερη πληροφορία σε σχέση με τα άλλα δύο.



Η πληροφορία έχει κεντρικό ρόλο στη ΘΑΣ καθώς συμβάλλει καθοριστικά στη διαμόρφωση του εργαλείου μέτρησης καθώς και στην τροποποίηση ήδη υφιστάμενων (Morizot et al., 2007. Toland, 2014). Για παράδειγμα, εάν σκοπός του ερευνητή είναι να διαχωρίζει άτομα με υψηλό επίπεδο λανθάνοντος χαρακτηριστικού (π.χ., αξιολόγηση κατάθλιψης στο πεδίο της κλινικής ψυχολογίας) τότε θα πρέπει να επιλεγούν εκείνα τα στοιχεία που να παρέχουν επαρκή πληροφορία στη συγκεκριμένη περιοχή. Αντιθέτως, εάν σκοπός είναι η μέτρηση του χαρακτηριστικού σε όλο το εύρος του, τότε θα πρέπει να επιλεγούν στοιχεία που να παρέχουν ικανοποιητική πληροφορία σε διαφορετικές περιοχές (π.χ., κάτω από τη μέση επίδοση, στην περιοχή της μέσης επίδοσης, και πάνω από τη μέση επίδοση). Η πληροφορία του κάθε στοιχείου μπορεί να αθροιστεί και να υπολογιστεί η συνολική πληροφορία που παρέχει το εργαλείο μέτρησης (Scale Information Function, SIF ή Test Information Function, TIF).

Επιπροσθέτως, η πληροφορία συμβάλλει στη βελτίωση υφιστάμενων οργάνων μέτρησης ή στη δημιουργία συντομευμένων εκδόσεών τους. Με βάση την εξέταση των γραφημάτων ο ερευνητής μπορεί να διαπιστώσει κενά στη μέτρηση του λανθάνοντος χαρακτηριστικού και να δρομολογηθεί η αναδιατύπωση υπαρχόντων στοιχείων ή η δημιουργία νέων με σκοπό την πιο ακριβή μέτρηση των συμμετεχόντων (Theodorakis, Tsigilis, Wann, Liannopoulos, & Emadi, 2016). Από την άλλη μεριά, ο εντοπισμός και η διαγραφή στοιχείων που προσφέρουν χαμηλή πληροφορία, είτε παρέχουν παρόμοια πληροφορία στην ίδια περιοχή του λανθάνοντος χαρακτηριστικού μπορεί να οδηγήσει σε πιο αποτελεσματικές και αξιοποιήσιμες συντομευμένες εκδόσεις. Αυτός είναι και ο λόγος που οι Embretson και Reise (2000) δηλώνουν, αναφερόμενοι στα πλεονεκτήματα της ΘΑΣ, ότι “Shorter tests can be more reliable than longer tests” (σελ. 18). Συνεπώς, ενώ κατά την ΚΘΜ η βελτίωση της αξιοπιστίας επιτυγχάνεται με την προσθήκη επιπλέον στοιχείων, ο ερευνητής που χρησιμοποιεί την ΘΑΣ προσθέτει προσεκτικά επιλεγμένα στοιχεία ανάλογα με το σκοπό του εργαλείου μέτρησης ή διαγράφει τα περιττά.

### **Προϋποθέσεις εφαρμογής της ΘΑΣ**

Η ΘΑΣ περιλαμβάνει διάφορα στατιστικά μοντέλα, τα οποία στηρίζονται σε συγκεκριμένες προϋποθέσεις. Συνεπώς, η ανάλυση εμπειρικών δεδομένων υπό το πρίσμα της ΘΑΣ μπορεί να παράσχει χρήσιμα αποτελέσματα αρκεί να ικανοποιούνται οι παρακάτω προϋποθέσεις: α) ύπαρξη ενός λανθάνοντος χαρακτηριστικού (unidimensionality), β) τοπική ανεξαρτησία (local independence) και γ) μονοτονία (monotonicity). Οι προϋποθέσεις αυτές παρουσιάζονται αναλυτικά παρακάτω.

### **Ύπαρξη ενός λανθάνοντος χαρακτηριστικού**

Μια βασική προϋπόθεση εφαρμογής της ΘΑΣ είναι ότι το εργαλείο μέτρησης αξιολογεί ένα μόνο λανθάνον χαρακτηριστικό. Η ικανοποίηση της προϋπόθεσης αυτής μπορεί να εξεταστεί με την εφαρμογή διερευνητικής παραγοντικής ανάλυσης (ΔΠΑ), στην οποία ο πίνακας τετραγωνικών ή πολυγωνικών συσχετίσεων θα πρέπει να εισαχθεί προς ανάλυση (Reid, Kolakowsky-Hayner, Lewis, & Armstrong, 2007). Βεβαίως, πολλές φορές στην πράξη τα εργαλεία μέτρησης αξιολογούν πάνω από μία διαστάσεις (Edwards, 2009. Price, 2017). Στην περίπτωση αυτή, ο ερευνητής έχει δύο επιλογές:

α) Εάν από την εφαρμογή της ΔΠΑ προκύψει ένας κυρίαρχος παράγοντας, με ουσιαστικά υψηλότερη ιδιοτιμή σε σχέση με τους υπόλοιπους, τότε το εργαλείο μπορεί να θεωρηθεί ως «επαρκώς μονοδιάστατο» (unidimensional enough) (Morizot et al., 2007) και η ανάλυση να προχωρήσει όπως είχε σχεδιαστεί. Η οπτική εξέταση του διαγράμματος ιδιοτιμών (scree-plot) είναι μία από τις προσεγγίσεις που έχουν αναφερθεί για τον έλεγχο του αριθμού των διαστάσεων. Μία άλλη συχνά χρησιμοποιούμενη προσέγγιση είναι η εξέταση της αναλογίας μεταξύ πρώτης και δεύτερης ιδιοτιμής. Για παράδειγμα, οι Morizot et al. (2007) αναφέρουν μία αναλογία 3 ή μεγαλύτερη ως ένδειξη μονοδιάστατου εργαλείου. Άλλοι ερευνητές στηρίζονται στο ποσοστό της διακύμανσης του πρώτου παράγοντα. Ο Reckase (1979) αναφέρει ότι όταν ο πρώτος παράγοντας εξηγεί τουλάχιστον το 20% της συνολικής διακύμανσης ενώ οι επόμενοι παράγοντες ερμηνεύουν ουσιαστικώς μικρότερα ποσοστά τότε το εργαλείο μπορεί να θεωρηθεί ως επαρκώς μονοδιάστατο.

Εάν για την εξέταση της ύπαρξης ενός λανθάνοντα παράγοντα χρησιμοποιηθεί η επιβεβαιωτική παραγοντική ανάλυση τότε καμία συσχέτιση μεταξύ των υπολοίπων της μέτρησης δε θα πρέπει να υπερβαίνει το .20 (Morizot et al., 2007). Η εξέταση ενός μοντέλου γενικού και ειδικών παραγόντων (bi-factor model) είναι μια άλλη ενδιαφέρουσα πρόταση (Reise & Haviland, 2005). Σύμφωνα με το μοντέλο αυτό, η μεταβλητότητα των στοιχείων ερμηνεύεται από δύο παράγοντες: ένα γενικό παράγοντα, ο οποίος είναι κοινός για όλα τα στοιχεία, και άλλους ειδικούς για κάθε διάσταση του εργαλείου μέτρησης (Morizot et al., 2007). Οι ειδικοί παράγοντες είναι ανεξάρτητοι μεταξύ τους καθώς και με το γενικό. Στην περίπτωση αυτή, ο ερευνητής μπορεί να υπολογίσει το ποσοστό της μεταβλητότητας που οφείλεται στον κύριο παράγοντα και το ποσοστό που οφείλεται στους δευτερεύοντες παράγοντες. Εάν βρεθεί ότι ο γενικός παράγοντας ερμηνεύει μεγαλύτερο μέρος της μεταβλητότητας σε σχέση με του ειδικούς τότε το εργαλείο μπορεί να χαρακτηριστεί ως μονοδιάστατο.

β) Στην περίπτωση που τα δεδομένα δεν υποστηρίζουν έναν κυρίαρχο παράγοντα μπορούν να εφαρμοστούν πιο σύνθετες στατιστικές τεχνικές που αναφέρονται ως Πολυδιάστατη Θεωρία Απόκρισης Στοιχείου (Multidimensional Item Response

Theory, MIRT). Υπάρχουν αρκετά προγράμματα, τόσο εμπορικά (π.χ., IRTPRO, flexMIRT, Mplus,) όσο και ανοικτού κώδικα (π.χ., βιβλιοθήκες mirt και TAM στο περιβάλλον R) που υποστηρίζουν αναλύσεις απόκρισης στοιχείου πολυδιάστατων εργαλείων μέτρησης (Chalmers, 2012. Han & Paek, 2014. Sims, 2017).

### **Τοπική ανεξαρτησία στοιχείων**

Μια άλλη προϋπόθεση που συνδέεται με την ύπαρξη ενός λανθάνοντος χαρακτηριστικού αφορά την τοπική ανεξαρτησία (local independence, LD) των στοιχείων. Αυτό σημαίνει ότι μετά την αφαίρεση της επίδρασης της λανθάνουσας ικανότητας, τα στοιχεία του εργαλείου μέτρησης αναμένεται να μη συνδέονται μεταξύ τους. Παραβίαση της τοπικής ανεξαρτησίας οδηγεί σε αύξηση των τιμών διάκρισης ( $\alpha$ ), που με τη σειρά της οδηγεί σε ψευδή επίπεδα ακρίβειας-πληροφορίας των μετρήσεων. Επιπλέον δίνεται η εσφαλμένη εικόνα ότι τα στοιχεία αυτά είναι οι καλύτεροι εκπρόσωποι του λανθάνοντος χαρακτηριστικού (Morizot et al., 2007. Toland, 2014).

Η ικανοποίηση της προϋπόθεσης αυτής εξετάζεται με τις τυποποιημένες  $\chi^2$  τιμές τοπικής ανεξαρτησίας για κάθε ζεύγος μεταβλητών (LD  $\chi^2$ ). Οι Cai et al. (2017) αναφέρουν ότι τιμές μικρότερες του |5| θεωρούνται μικρές και αμελητέες, τιμές ανάμεσα στο |5| και |10| θεωρούνται μέτριες, ενώ τιμές μεγαλύτερες από το |10| είναι υψηλές και πιθανόν να δηλώνουν προβλήματα τοπικής εξάρτησης. Επειδή όμως η απουσία ή ο μικρός αριθμός παρατηρήσεων σε κάποιο κελί μπορεί να δημιουργήσει ζητήματα τοπικής εξάρτησης προτείνεται η ερμηνεία της τιμής LD  $\chi^2$  να συνοδεύεται και από την εξέταση του πίνακα συνάφειας (Toland, 2014). Τέλος, εάν υπάρχει ζήτημα τοπικής εξάρτησης θα πρέπει να εξεταστεί η επίδρασή της στην εκτίμηση των παραμέτρων και στην προσαρμογή των στοιχείων. Δηλαδή να μελετηθεί ο βαθμός διαφοροποίησης τόσο των παραμέτρων όσο και της προσαρμογής των στοιχείων με την εισαγωγή και τον αποκλεισμό από την ανάλυση των στοιχείων που είναι ύποπτα για δημιουργία τοπικής εξάρτησης (Edelen & Reeve, 2007).

### **Μονοτονία**

Η προϋπόθεση αυτή στηρίζεται στη μονοτονικά αυξανόσα λογαριθμική συνάρτηση (monotonic logistic function) μεταξύ του λανθάνοντος χαρακτηριστικού και της πιθανότητας ορθής επιλογής. Αυτό σημαίνει ότι καθώς το λανθάνον χαρακτηριστικό αυξάνεται, η πιθανότητα ορθής επιλογής επίσης αυξάνεται. Στην περίπτωση που τα δεδομένα δεν ακολουθούν το μοντέλο αυτό, η χρήση τους και η ανάλυση δεν έχει ιδιαίτερη χρησιμότητα. Η διερεύνηση της ικανοποίησης αυτής της προϋπόθεσης πραγματοποιείται με το στατιστικό S-X<sup>2</sup> (Orlando & Thissen, 2000, 2003). Το στατιστικό

αυτό είναι μια παραλλαγή του Pearson  $\chi^2$  για την εξέταση της προσαρμογής των παρατηρούμενων αποκρίσεων του στοιχείου με τις αναμενόμενες αποκρίσεις σύμφωνα με το μοντέλο που έχει επιλεγεί (π.χ., 1PL, 2PL) και εφαρμόζεται για κάθε στοιχείο του εργαλείου μέτρησης.

### Επιλογή του κατάλληλου μοντέλου

Μετά την εξέταση των προϋποθέσεων της ανάλυσης ακολουθεί η επιλογή του πιο κατάλληλου μοντέλου για τα συγκεκριμένα κατά περίπτωση εμπειρικά δεδομένα. Η χρησιμότητα της ΘΑΣ και η υπεροχή της απέναντι στην ΚΘΜ έχει νόημα μόνο όταν το μοντέλο που θα επιλεγεί έχει καλή προσαρμογή στα δεδομένα. Έτσι, ο ερευνητής καλείται να επιλέξει μεταξύ του πιο φειδωλού μοντέλου (π.χ., 1PL ή 2PL) που όμως μπορεί να μην έχει καλή προσαρμογή στα δεδομένα, και ενός πιο σύνθετου (π.χ., 2PL ή 3PL) με καλύτερη ίσως προσαρμογή.

Η επιλογή του μοντέλου που περιγράφει καλύτερα τα δεδομένα γίνεται σε δύο επίπεδα. Το πρώτο επίπεδο αφορά τη συνολική προσαρμογή του εργαλείου μέτρησης και το δεύτερο την προσαρμογή του κάθε στοιχείου ξεχωριστά. Ο ερευνητής θα πρέπει να λάβει υπόψη του τα αποτελέσματα των δύο αυτών αναλύσεων προκειμένου να τεκμηριώσει την τελική επιλογή του.

Σε επίπεδο συνολικής προσαρμογής το πρόγραμμα IRTPRO, καθώς και άλλα προγράμματα, εμφανίζει τις τιμές του λογαρίθμου πιθανοφάνειας (log likelihood), του Akaike Information Criterion (AIC, Akaike, 1974) και του Bayesian Information Criterion (Schwarz, 1978). Χαμηλές τιμές των δεικτών αυτών δηλώνουν καλύτερη προσαρμογή. Δυστυχώς, δεν υπάρχουν συγκεκριμένες τιμές του κάθε κριτηρίου με βάση τις οποίες να τεκμαίρεται η προσαρμογή του μοντέλου. Συνεπώς, η χρησιμότητα των δεικτών αυτών περιορίζεται στη σύγκριση διάφορων υποψήφιας προς υιοθέτηση μοντέλων. Ο πρώτος δείκτης (log likelihood) είναι χρήσιμος για τη σύγκριση δύο εμφωλευμένων μοντέλων (nested models). Εάν η τιμή του πολλαπλασιαστεί με το -2 τότε ακολουθεί την κατανομή  $\chi^2$ . Στην περίπτωση που δύο υποψήφια μοντέλα είναι εμφωλευμένα (π.χ., 1PL και 2PL), τότε η διαφορά των τιμών των  $\chi^2$  ακολουθεί και αυτή την  $\chi^2$  κατανομή με βαθμούς ελευθερίας τη διαφορά των βαθμών ελευθερίας των δύο μοντέλων. Μη στατιστικά σημαντικές τιμές δηλώνουν ότι τα δύο μοντέλα δε διαφέρουν και θα πρέπει να επιλεγεί το πιο φειδωλό. Στην αντίθετη περίπτωση, όταν δηλαδή η τιμή του  $\Delta\chi^2$  είναι στατιστικά σημαντική, τα δύο μοντέλα διαφέρουν ως προς την προσαρμογή τους και επιλέγεται αυτό με τις χαμηλότερες τιμές. Οι δείκτες AIC και BIC μπορεί να λειτουργήσουν επικουρικά στην επιλογή του μοντέλου. Η σημασία τους όμως είναι πιο ουσιαστική στην περίπτωση μη-εμφωλευμένων μοντέλων.

Επιπροσθέτως, το IRTPRO μπορεί να εμφανίσει (εάν ζητηθεί) το δείκτη  $M_2$  (Maydeu-

Olivares & Joe, 2005, 2006) που έχει πρόσφατα αναπτυχθεί για την αξιολόγηση της συνολικής προσαρμογής του εργαλείου μέτρησης. Το πλεονέκτημα του δείκτη αυτού είναι ότι συνοδεύεται από  $p$ -τιμές και τον αντίστοιχο δείκτη RMSEA. Στην περίπτωση αυτή, ο ερευνητής έχει ένα σαφές κριτήριο (π.χ.,  $p$ -τιμή < .05) με βάση το οποίο να συναγάγει συμπέρασμα σχετικά με τη συνολική προσαρμογή του συγκεκριμένου μοντέλου, χωρίς να απαιτείται η σύγκρισή του με κάποιο άλλο.

Παράλληλα με τη διερεύνηση του πλέον κατάλληλου μοντέλου, εξετάζεται και η προσαρμογή του κάθε στοιχείου. Στην περίπτωση χρήσης του στατιστικού προγράμματος IRTPRO αυτή πραγματοποιείται με το στατιστικό  $S-X^2$  όπως έχει αναφερθεί παραπάνω. Ανάλογα με τα αποτελέσματα της ανάλυσης μπορεί να απομακρυνθούν στοιχεία που αποκλίνουν σημαντικά από το συγκεκριμένο μοντέλο (π.χ., 2PL). Παρά το γεγονός ότι η προσαρμογή του κάθε στοιχείου εξετάζεται σε επίπεδο σημαντικότητας .05 προτείνεται να γίνεται κάποια διόρθωση (π.χ., Bonferroni, Holm-Bonferroni, Benjamini-Hochberg) λόγω του πολλαπλού αριθμού στοιχείων που εξετάζονται (Magis, Béland, Tuerlinckx, & de Boeck, 2010. Nguyen, Han, Kim, & Chan, 2014).

### Παράδειγμα εφαρμογής της Θεωρίας Απόκρισης Στοιχείου

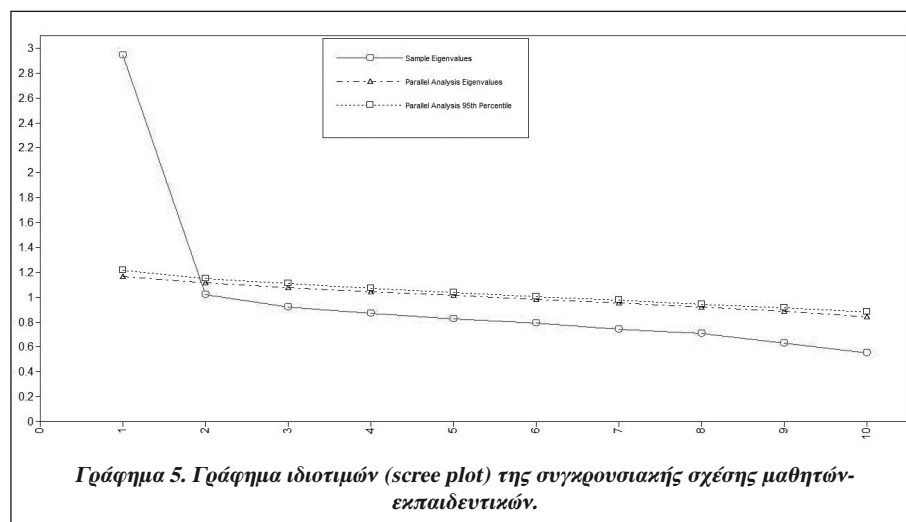
Στο τμήμα αυτό της εργασίας θα παρουσιαστεί ένα παράδειγμα εφαρμογής της ΘΑΣ με σκοπό οι έννοιες που παρουσιάστηκαν παραπάνω να γίνουν πιο κατανοητές. Τα δεδομένα προέρχονται από 1904 μαθητές προσχολικής ηλικίας που απάντησαν το Young Children's Appraisals of Teacher Support (Y-CATS) (Mantzicopoulos & Neuharth-Pritchett, 2003). Το Y-CATS περιέχει 27 δηλώσεις (στοιχεία) για την αξιολόγηση τριών διαστάσεων των αντιλήψεων των μαθητών για τις σχέσεις τους με τον εκπαιδευτικό: εγγύτητα (12 δηλώσεις), σύγκρουση (10 δηλώσεις) και αυτονομία (5 δηλώσεις). Οι απαντήσεις δίνονται σε δίτιμη κλίμακα: Συμφωνία με τη δήλωση – Διαφωνία με τη δήλωση. Για τις ανάγκες της παρούσας εργασίας η ανάλυση θα περιοριστεί στην ανάλυση των αντιλήψεων των μαθητών για τη συγκρουσιακή τους σχέση με τον εκπαιδευτικό και σε 900 περιπτώσεις. Με τον όρο συγκρουσιακή σχέση νοείται η αντιλαμβανόμενη αρνητικότητα και αντιπαράθεση στη σχέση μεταξύ μαθητή και εκπαιδευτικού. Δύο ενδεικτικά στοιχεία είναι: «Ο δάσκαλός μου είναι κακός» και «Ο δάσκαλός μου λέει ότι κάνω κάτι λάθος συχνά».

Η ανάλυση θα πραγματοποιηθεί με τη φοιτητική έκδοση του προγράμματος IRTPRO ver. 4.2 (Cai et al., 2017). Το πρόγραμμα αυτό επιλέχθηκε επειδή έχει γραφικό περιβάλλον, είναι αρκετά φιλικό προς το χρήστη και μπορεί να φιλοξενήσει πολλά είδη της ΘΑΣ (Han & Paek, 2014. Paek & Han, 2012). Επιπλέον, η φοιτητική έκδοση διαθέτει όλες τις δυνατότητες ανάλυσης της πλήρους έκδοσης με τον περιορι-

σμό ότι μπορεί να δεχθεί προς ανάλυση 25 μεταβλητές για 1000 συμμετέχοντες. Θεωρούμε ότι για αρκετές μελέτες στο χώρο των κοινωνικών επιστημών η φοιτητική έκδοση είναι ικανοποιητική. Στα Παράρτημα Α υπάρχουν τα αποτελέσματα των ανάλυσεων με το πρόγραμμα IRTPRO. Στο Παράρτημα Β περιγράφεται η ίδια ανάλυση αλλά με τη χρήση του προγράμματος ανοικτού κώδικα ltm (έκδοση 1.1-1, στις 17-4-2018) στο περιβάλλον της R.

### Βήμα 1. Εξέταση της παραγοντικής δομής

Στο βήμα αυτό ο ερευνητής καλείται να διερευνήσει τον αριθμό των λανθανουσών μεταβλητών του προς εξέταση οργάνου μέτρησης. Για την εξέταση της παραγοντικής δομής των 10 στοιχείων της συγκρουσιακής σχέσης εφαρμόστηκε διερευνητική ανάλυση παραγόντων. Από την ανάλυση προέκυψαν μικτά συμπεράσματα. Πιο συγκεκριμένα, αρχικά εμφανίστηκαν δύο παράγοντες με ιδιοτιμές πάνω από τη μονάδα (2.947 και 1.019 αντιστοίχως). Η αναλογία τους (2.89) ήταν οριακά μικρότερη από αυτήν που προτείνουν διάφοροι ερευνητές για να χαρακτηριστεί το εργαλείο ως επαρκώς μονοδιάστατο. Το διάγραμμα ιδιοτιμών, όμως, καθώς και η παράλληλη ανάλυση (με αριθμό επαναλήψεων 1000) δείχνουν καθαρά την ύπαρξη ενός παράγοντα (Γράφημα 5). Με δεδομένο ότι η παράλληλη ανάλυση έχει χαρακτηριστεί ως μία αρκετά αξιόπιστη μέθοδος εκτίμησης του αριθμού των παραγόντων που πρέπει να διατηρηθούν (Hayton, Allen, & Scarpello, 2004. Lance, Butts, & Michels, 2006) μπορούμε να δεχθούμε την ύπαρξη μίας διάστασης και, συνεπώς, η ανάλυση της ΘΑΣ μπορεί να προχωρήσει χρησιμοποιώντας τις τεχνικές της μονοδιάστατης ανάλυσης.



## Βήμα 2. Επιλογή μεταξύ υποψήφιων μοντέλων

Στο βήμα αυτό εξετάζεται ποιο είναι το καλύτερο μοντέλο για να περιγράψει τις απαντήσεις των συμμετεχόντων στο Y-CATS. Η μελέτη της προσαρμογής γίνεται σε δύο επίπεδα: σε επίπεδο προσαρμογής του εργαλείου συνολικά και σε επίπεδο προσαρμογής του κάθε στοιχείου. Στο συγκεκριμένο παράδειγμα θα εξεταστούν δύο μοντέλα: α) ένα μοντέλο μίας παραμέτρου (1PL), στο οποίο όλα τα στοιχεία έχουν κοινή τιμή διάκρισης για όλα τα στοιχεία και διαφορετικές τιμές δυσκολίας (11 βαθμοί ελευθερίας) και β) ένα μοντέλο δύο παραμέτρων (2PL), στο οποίο τόσο η διάκριση όσο και η δυσκολία διαφοροποιούνται ανά στοιχείο (20 βαθμοί ελευθερίας).

Στον Πίνακα 1 παρουσιάζονται οι δείκτες προσαρμογής που παρέχει το πρόγραμμα IRTPRO 4.2. Από τις τιμές των δεικτών αυτών είναι προφανές ότι το μοντέλο των δύο παραμέτρων έχει καλύτερη προσαρμογή στα δεδομένα (π.χ., AIC, RMSEA). Επιπλέον, μπορεί να γίνει σύγκριση των δύο αυτών μοντέλων καθώς είναι εμφωλευμένα. Εκτέλεση του Ελέγχου του Λόγου Πιθανοφαιγιών (Likelihood Ratio Test) έδειξε στατιστικώς σημαντική διαφορά ( $\Delta\chi^2 = 77.02$ ,  $\Delta df = 9$ ,  $p < .001$ )<sup>5</sup>, συνεπώς το μοντέλο των δύο παραμέτρων φαίνεται να έχει συνολικά καλύτερη προσαρμογή σε σχέση με το μοντέλο της μιας παραμέτρου.

**Πίνακας 1. Δείκτες προσαρμογής της συγκρουσιακής σχέσης του Y-CATS για τα δύο υποψήφια μοντέλα f**

	1PL	2PL
-2 Loglikelihood	10516.15 ( $df = 11$ )	10439.13 ( $df = 20$ )
AIC	10538.15	10479.13
BIC	10590.97	10575.18
M2	150.05 ( $df = 44$ )	69.88 ( $df = 35$ )
RMSEA	.05	.03

**Σημείωση:** 1PL = μοντέλο μίας παραμέτρου (δυσκολίας), 2PL = μοντέλο δύο παραμέτρων (διάκρισης και δυσκολίας)

<sup>5</sup>. Το μοντέλο των τριών παραμέτρων [trm()] δεν εμφάνισε στατιστικά καλύτερη προσαρμογή σε σχέση με το μοντέλο των δύο παραμέτρων > anova(conflict\_2PL, conflict\_3PL)

Likelihood Ratio Table

	AIC	BIC	log.Lik	LRT	df	p.value
conflict_2PL	10479.12	10575.17	-5219.56			
conflict_3PL	10497.81	10641.89	-5218.91	1.31	10	0.999



Στη συνέχεια εξετάζεται η προσαρμογή του κάθε στοιχείου ξεχωριστά με το δείκτη  $S-X^2$ . Από τα αποτελέσματα φαίνεται ότι τα στοιχεία της συγκρουσιακής σχέσης έχουν καλύτερη προσαρμογή στο μοντέλο των δύο παραμέτρων (Πίνακας 2). Πιο αναλυτικά, όσον αφορά το μοντέλο της μίας παραμέτρου τέσσερα από τα δέκα στοιχεία έχουν στατιστικώς σημαντικές τιμές (y14co, y17co, y20co και y22co), γεγονός που δηλώνει έλλειψη καλής προσαρμογής. Από την άλλη μεριά, το μοντέλο των δύο παραμέτρων έχει συστηματικά χαμηλότερες τιμές  $S-X^2$  σε σύγκριση με αυτό της μίας παραμέτρου και επιπλέον καμία τιμή του δεν εμφανίζεται ως στατιστικώς σημαντική.

**Πίνακας 2. Δείκτες προσαρμογής του κάθε στοιχείου της συγκρουσιακής σχέσης του Y-CATS για τα δύο υποψήφια μοντέλα**

	1PL			2PL		
	S-X <sup>2</sup>	df	p-value	S-X <sup>2</sup>	df	p-value
y2co	11.83	8	0.1584	6.43	8	0.6007
y5co	9.13	8	0.3331	5.08	8	0.7501
y8co	11.93	8	0.1541	10.84	8	0.2106
y11co	9.95	8	0.2702	10.09	8	0.2584
y14co	18.23	8	0.0195	8.94	8	0.3491
y17co	27.99	8	0.0005	11.31	8	0.1845
y20co	18.30	8	0.0191	12.09	8	0.1471
y22co	34.75	7	0.0001	8.10	8	0.4251
y24co	12.93	8	0.1138	6.16	8	0.6301
y26co	10.81	8	0.2122	11.85	8	0.1574

**Σημείωση:** 1PL = μοντέλο μίας παραμέτρου (δυσκολίας), 2PL = μοντέλο δύο παραμέτρων (διάκρισης και δυσκολίας)

Τέλος, ένα άλλο στοιχείο που επίσης συνηγορεί υπέρ του μοντέλου των δύο παραμέτρων είναι τα αποτελέσματα της εξέτασης της τοπικής ανεξαρτησίας (βλέπε Παράρτημα Α - Marginal fit (X<sup>2</sup>) and Standardized LD X<sup>2</sup> Statistics for Group 1). Η μεγαλύτερη τιμή που εμφανίζεται μετά την αφαίρεση της επίδρασης της λανθάνουσας μεταβλητής είναι 3.1, η οποία είναι σαφώς μικρότερη του |5| που προτείνουν οι Cai et al. (2017). Αντιθέτως, στο μοντέλο της μίας παραμέτρου εμφανίζονται τέσσερεις τιμές πάνω από |10| και μία μεγαλύτερη του |5| (Παράρτημα Α). Με βάση τα παραπάνω αποτελέσματα μπορούμε να δεχθούμε με ασφάλεια ότι το μοντέλο δυο παραμέτρων περιγράφει αρκετά ικανοποιητικά τις απαντήσεις των μαθητών στις δέκα δηλώσεις συγκρουσιακής σχέσης με τον εκπαιδευτικό και μπορούμε να προχωρήσουμε στην παρουσίαση και την ερμηνεία των παραμέτρων.

### **Βήμα 3. Παρουσίαση και ερμηνεία των παραμέτρων**

Στον Πίνακα 3 παρουσιάζονται οι τιμές των δύο παραμέτρων διάκρισης και δυσκολίας για κάθε δήλωση μαζί με τα τυπικά τους σφάλματα. Οι περισσότερες τιμές της διάκρισης είναι μεγαλύτερες της μονάδας, αποτέλεσμα που δείχνει ότι τα στοιχεία μπορεί να ξεχωρίσουν συμμετέχοντες με παραπλήσια επίπεδα αντιλαμβανόμενης συγκρουσιακής σχέσης με τον εκπαιδευτικό. Η χαμηλότερη τιμή εμφανίζεται για το στοιχείο y22co που σύμφωνα με τον Baker (2001, σελ. 34) θεωρείται μέτρια. Όσον αφορά τη δυσκολία των στοιχείων, αυτή φαίνεται να κυμαίνεται σε επίπεδα μεγαλύτερα της μέσης αντιλαμβανόμενης συγκρουσιακής σχέσης, αφού οι περισσότερες τιμές είναι θετικές. Αυτό σημαίνει ότι εννέα από τα δέκα στοιχεία αξιολογούν το λανθάνον χαρακτηριστικό πάνω από το μεσαίο επίπεδο. Ένα μόνο στοιχείο (y22co) έχει αρνητική τιμή που σημαίνει ότι είναι χρήσιμο για την αξιολόγηση χαμηλών επιπέδων συγκρουσιακής σχέσης.

Αξίζει να τονιστεί ότι επτά από τα δέκα στοιχεία αξιολογούν το λανθάνον χαρακτηριστικό σε ένα πολύ μικρό εύρος ( $.76 - .35 = .41$ ), που είναι μικρότερο από μισή τυπική απόκλιση. Επιπλέον, τέσσερα από αυτά τα στοιχεία έχουν σχεδόν παρόμοιες τιμές δυσκολίας (y11co και y20co καθώς και y8co και y26co). Τα ευρήματα αυτά δείχνουν με ξεκάθαρο τρόπο ότι τα δέκα στοιχεία της συγκρουσιακής σχέσης του εργαλείου Y-CATS δεν αξιολογούν όλο το εύρος του λανθάνοντος χαρακτηριστικού και ότι αρκετά από αυτά είναι περιττά και θα μπορούσαν να αποκλειστούν από μελλοντικές μελέτες. Το αποτέλεσμα αυτό είναι αρκετά σημαντικό για την εγκυρότητα της μέτρησης της συγκρουσιακής σχέσης και δύσκολα θα μπορούσε να συναχθεί πα-

**Πίνακας 3. Τιμές διάκρισης και δυσκολίας για το μοντέλο των δύο παραμέτρων**

	<i>a</i>	<i>se</i>	<i>c</i>	<i>se</i>	<i>b</i>	<i>se</i>
y2co	1.70	0.19	-1.90	0.15	1.12	0.09
y5co	0.98	0.11	-0.49	0.08	0.50	0.09
y8co	1.15	0.12	-0.41	0.08	0.35	0.07
y11co	1.19	0.13	-0.54	0.09	0.45	0.08
y14co	1.79	0.19	-1.36	0.13	0.76	0.07
y17co	0.78	0.10	-0.82	0.08	1.05	0.15
y20co	1.61	0.17	-0.74	0.10	0.46	0.06
y22co	0.66	0.10	1.04	0.08	-1.58	0.25
y24co	1.69	0.18	-1.00	0.11	0.59	0.06
y26co	1.23	0.13	-0.47	0.09	0.38	0.07

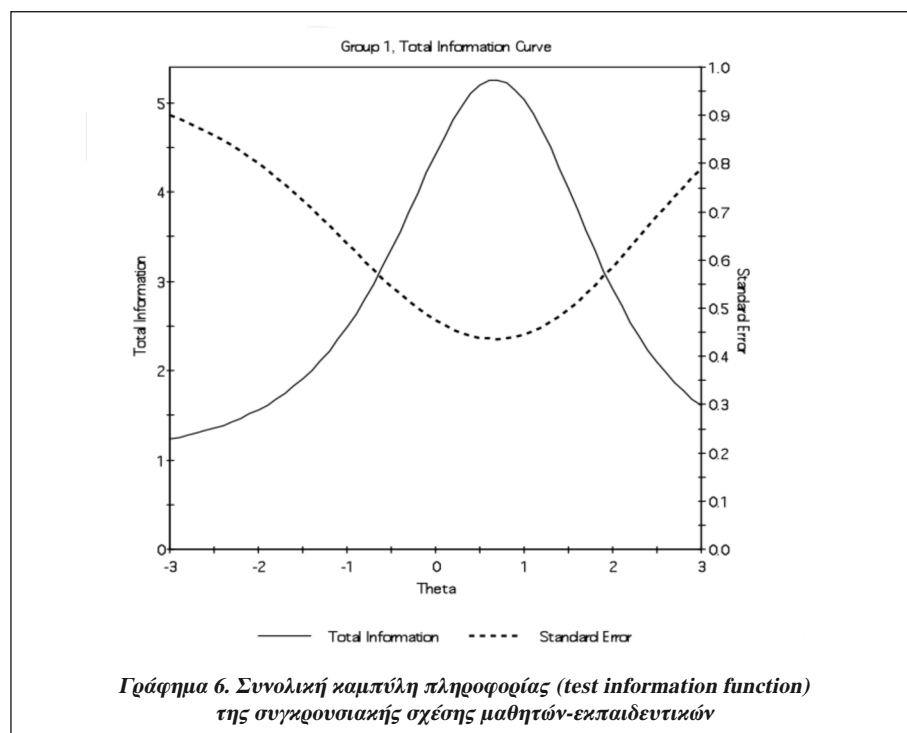
**Σημείωση:** *a* = παράμετρος διάκρισης, *c* = σταθερός όρος (intercept), *b* = παράμετρος δυσκολίας, *se* = τυπικό σφάλμα.  $b = -c/a$

ρόμοιο συμπέρασμα με τη χρήση των μεθόδων της ΚΘΜ. Επίσης, δείχνει την κατεύθυνση στην οποία θα πρέπει να κινηθεί ο ερευνητής για τη βελτίωση της μέτρησης και την καλύτερη αποτύπωση της συγκρουσιακής σχέσης.

#### Βήμα 4. Μελέτη της πληροφορίας

Η μελέτη της πληροφορίας που παρέχουν οι δηλώσεις της συγκρουσιακής σχέσης των μαθητών με τους εκπαιδευτικούς θα εξεταστεί σε δύο επίπεδα: α) σε επίπεδο στοιχείου και β) σε συνολικό επίπεδο. Στο Παράρτημα Α παρουσιάζεται η πληροφορία για κάθε ένα από τα δέκα στοιχεία. Από την οπτική εξέταση των γραφημάτων αβίαστα προκύπτει το συμπέρασμα ότι τα στοιχεία  $y_{2co}$ ,  $y_{14co}$ ,  $y_{20co}$  και  $y_{24co}$  διαθέτουν το μεγαλύτερο ποσό πληροφορίας, σε αντίθεση με τα στοιχεία  $y_{17co}$  και  $y_{22co}$ . Συνεπώς, εάν ο ερευνητής επιθυμεί να ελαττώσει τον αριθμό των στοιχείων χωρίς ουσιαστική μείωση της πληροφορίας τότε είναι προφανές ότι θα πρέπει να απομακρύνει τα δύο τελευταία στοιχεία.

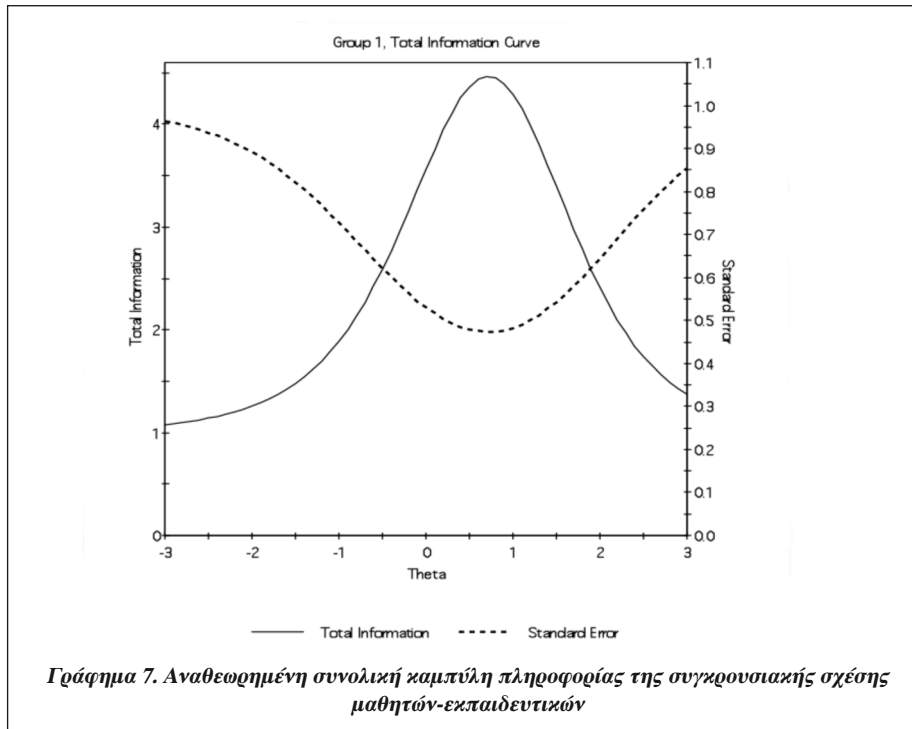
Στο Γράφημα 6 παρουσιάζεται το συνολικό ποσό πληροφορίας που παρέχουν οι δέκα δηλώσεις. Με διακεκομμένη γραμμή παριστάνεται το τυπικό σφάλμα. Εύκολα



διαπιστώνεται ότι οι δύο αυτές έννοιες είναι αντιστρόφως ανάλογες. Δηλαδή, όπου παρατηρείται αύξηση της πληροφορίας το τυπικό σφάλμα μειώνεται και αντιστρόφως. Το λανθάνον χαρακτηριστικό της συγκρουσιακής σχέσης μετρείται με τη μεγαλύτερη ακρίβεια περίπου 0.75 τυπικές αποκλίσεις πάνω από το μέσο επίπεδό της. Επιπλέον, οι δέκα δηλώσεις μπορεί να χρησιμοποιηθούν για να μετρήσουν αξιόπιστα το λανθάνον χαρακτηριστικό από μέσο επίπεδο συγκρουσιακής σχέσης έως και 1.5 τυπική απόκλιση πάνω από αυτό. Έξω από αυτά τα όρια, και καθώς απομακρυνόμαστε, η αξιοπιστία της μέτρησης μειώνεται αισθητά.

### **Βήμα 5. Αναθεώρηση της μέτρησης της συγκρουσιακής σχέσης**

Με βάση τα παραπάνω ευρήματα ο ερευνητής θα μπορούσε να βελτιώσει τη μέτρηση της συγκρουσιακής σχέσης με δύο τρόπους, οι οποίοι μπορεί να είναι συμπληρωματικοί. Αρχικά θα μπορούσε να απομακρύνει στοιχεία που: α) είναι περιττά καθώς αξιολογούν παρόμοια επίπεδα του λανθάνοντος χαρακτηριστικού (y8co και y11co) και β) δε συνεισφέρουν στην πληροφορία (y17co και y22co). Διαγραφή αυτών των στοιχείων από την ανάλυση και υπολογισμός εκ νέου της πληροφορίας από τα ενα-



πομείναντα έξι στοιχεία δείχνει παρόμοιο επίπεδο πληροφορίας με αυτό των δέκα στοιχείων στο ίδιο εύρος τιμών (Γράφημα 7). Μία άλλη επιλογή, αρκετά πιο επίπονη αλλά και πιο ουσιαστική είναι η ανάπτυξη νέων στοιχείων, τα οποία να αξιολογούν τόσο χαμηλά ( $< 0$ ) όσο και αρκετά υψηλά ( $> 1.5$ ) επίπεδα της αντιλαμβανόμενης συγκρουσιακής σχέσης των μαθητών με τον εκπαιδευτικό, με την προϋπόθεση ότι ο ερευνητής ενδιαφέρεται να αξιολογήσει τη συγκρουσιακή σχέση σε όλο το μήκος του λανθάνοντος χαρακτηριστικού.

### **Βήμα 6. Υπολογισμός τιμών λανθάνοντος χαρακτηριστικού**

Το τελευταίο βήμα της ανάλυσης αφορά τον υπολογισμό των τιμών του λανθάνοντος χαρακτηριστικού για κάθε συμμετέχοντα. Ο υπολογισμός στηρίζεται στο μοτίβο των αποκρίσεων σε συνδυασμό με τις παραμέτρους του κάθε στοιχείου (Edwards, 2009. Toland, 2014). Στη σχετική βιβλιογραφία υπάρχουν διάφοροι μέθοδοι υπολογισμού των τιμών του λανθάνοντος χαρακτηριστικού όπως η μέθοδος της Μέγιστης Πιθανοφάνειας (Maximum Likelihood, ML), η μέθοδος της μεγιστοποίησης της εκ των υστέρων πιθανότητας (Maximum a Posteriori, MAP), και η μέθοδος της εκ των υστέρων αναμενόμενης πιθανότητας (Expected a Posteriori, EAP). Οι Embretson και Reise (2000) παρουσιάζουν με αναλυτικό τρόπο, χρησιμοποιώντας συγκεκριμένο παράδειγμα, τις διάφορες μεθόδους υπολογισμού του λανθάνοντος χαρακτηριστικού.

Παρά το γεγονός ότι η μέθοδος ML είναι ευρέως γνωστή και συχνά η προεπιλεγμένη σε αρκετά στατιστικά προγράμματα, στο χώρο της ΘΑΣ έχει τα παρακάτω μειονεκτήματα: α) δεν μπορεί να εκτιμήσει το λανθάνον χαρακτηριστικό για άτομα με απόλυτη επιτυχία σε όλα τα στοιχεία (1) ή απόλυτη αποτυχία (0) και β) για να δώσει αξιόπιστα αποτελέσματα ως προς το λανθάνον χαρακτηριστικό χρειάζονται δοκιμασίες με μεγάλο αριθμό στοιχείων (π.χ., 50 στοιχεία) (Embretson & Reise, 2000). Από την άλλη μεριά, οι μέθοδοι MAP και EAP είναι Bayesians προσεγγίσεις, μπορούν να εφαρμοστούν για κάθε μοτίβο απόκρισης αλλά έχουν την τάση για μεροληψία ως προς τη μέση τιμή του πληθυσμού (Cai et al., 2017. Embretson & Reise, 2000). Τέλος, η μέθοδος EAP φαίνεται να υπερερεεί της MAP αφού εμφανίζει μικρότερο μέσο σφάλμα (Cai et al., 2017) και είναι η προεπιλεγμένη μέθοδος σε διάφορες στατιστικές εφαρμογές όπως το IRTPRO και mirt.

Ανεξάρτητα όμως από τη μέθοδο υπολογισμού που θα επιλεγεί, θα πρέπει να δοθεί μια μετρική στο λανθάνον χαρακτηριστικό. Η συνηθέστερη επιλογή είναι να τεθεί η μέση τιμή στο μηδέν και η τυπική απόκλιση στη μονάδα (Edwards, 2009. Embretson & Reise, 2000. Toland, 2014). Από την καρτέλα scoring του IRTPRO φαίνονται οι δύο αυτές προεπιλεγμένες ρυθμίσεις της μετρικής του λανθάνοντος χαρακτηριστικού καθώς και η επιλογή της EAP μεθόδου. Στην περίπτωση αυτή, δημιουργείται και ένα επι-

πλέον αρχείο της μορφής \*-sco.txt (στο συγκεκριμένο παράδειγμα “YCATS conflict. 2PL-sco.txt”) που περιέχει τέσσερεις στήλες: τη διάσταση (στην περίπτωση που έχει χρησιμοποιηθεί πολυδιάστατη ανάλυση), τον αύξοντα αριθμό του συμμετέχοντα, την εκτίμηση του λανθάνοντος χαρακτηριστικού και το τυπικό σφάλμα της εκτίμησης.

Οι τιμές του λανθάνοντος χαρακτηριστικού με βάση τα δέκα στοιχεία της συγκρουσιακής σχέσης κυμαίνονται σε ένα σχετικά περιορισμένο εύρος τιμών από -1.392 έως 2.007 ( $M.T. = 0.0001$ ,  $T.A. = .8461$ ). Το εύρος των τιμών μπορεί να συγκριθεί με το γράφημα της συνολικής πληροφορίας του εργαλείου έτσι ώστε ο ερευνητής να γνωρίζει για ποιους και πόσους συμμετέχοντες τα επίπεδα «συγκρουσιακής σχέσης του μαθητή με τον εκπαιδευτικό» έχουν υπολογιστεί αξιόπιστα. Επιπλέον, εκτίμηση του λανθάνοντος χαρακτηριστικού εκτός των συγκεκριμένων ορίων απαιτεί επιπρόσθετη, ειδικά στοχευμένη διερευνητική δραστηριότητα όπως αναφέρθηκε παραπάνω στο Βήμα 5. Τέλος, η νέα αυτή μεταβλητή μπορεί να χρησιμοποιηθεί σε περαιτέρω αναλύσεις όπως για την εξέταση άλλων ειδών εγκυρότητας της κλίμακας (π.χ., εγκυρότητα κριτηρίου) ή τη διαφοροποίησή της ανάλογα με επιλεγμένα δημογραφικά ή άλλα χαρακτηριστικά.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Baker, F. B. (2001). *The basis of item response theory* (2nd ed). ERIC Clearinghouse on Assessment and Evaluation, College Park. <http://edres.org/irt/>
- Cai, L., Thissen, D., & du Toit, S. H. C. (2017). *IRTPRO 4.2 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Chalmers, P. R. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. New York, NY: Guilford.
- DeMars, C. (2010). *Item response theory*. New York, NY: Oxford University Press.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(1), 5-18.
- Edwards, M. C. (2009). An introduction to Item Response Theory using the Need for Cognition Scale. *Social and Personality Psychology Compass*, 3(4), 507-529. doi.org/10.1111/j.1751-9004.2009.00194.x
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 253-262.
- Han K. T., & Paek, I. (2014). A review of commercial software packages for multidimensional IRT modeling. *Applied Psychological Measurement*, 38(6), 486 - 498. doi: 10.1177/0146621614536770
- Hayton, J. A., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7, 191-205.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, 9, 202-220.
- Magis, D., Béland, S., Tuerlinckx, F., & de Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847-862. doi:10.3758/BRM.42.3.847
- Magno, C. (2009). Demonstrating the difference between Classical Test Theory and Item Response Theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1-11.
- Mantzicopoulos, P., & Neuharth-Pritchett, S. (2003). Development and validation of a measure to assess head start children's appraisals of teacher support. *Journal of School Psychology*, 41, 431-451. doi:10.1016/j.jsp.2003.08.002
- Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009-1020.
- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in contingency tables. *Psychometrika*, 71, 713-732
- Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407-423). New York, NY: Guilford.
- Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient*, 7(1), 23-35. doi:10.1007/s40271-013-0041-0
- Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27, 289-298.
- Paek, I., & Han, K. T. (2012). IRTPRO 2.1 for Windows (Item Response Theory for Patient-Reported Outcomes). *Applied Psychological Measurement*, 37(3), 242-252. doi: 10.1177/0146621612468223
- Price, L. R. (2017). *Psychometric methods: Theory into practice*. New York, NY: Guilford.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, 17(5), 1-25.

- Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, A. N., & Armstrong, A. J. (2007). Modern psychometric methodology: Applications of Item Response Theory. *Rehabilitation Counseling Bulletin, 50*(3), 177-188.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207-230.
- Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment, 84*(3), 228-238.
- Sansone, R. A., Wiederman, M. W., & Sansone, L. (1998). The Self-Harm Inventory (SHI): Development of a scale for identifying self-destructive behaviors and borderline personality disorder. *Journal of Clinical Psychology, 54*, 973-983.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461 - 464.
- Sims, T. (2017). *Comparison of IRTPRO 3 and Mplus 7 for multidimensional item response item parameter and examinee ability estimation*. Doctoral Dissertation, Georgia State University. Retrieved from: [https://scholarworks.gsu.edu/eps\\_diss/186](https://scholarworks.gsu.edu/eps_diss/186)
- Theodorakis, N., Tsigilis, N., Wann, D., Liannopoulos, G., & Emadi, A. A. (2016). Sport Spectator Identification Scale: An item response analysis approach. *International Journal of Sport Management, 17*, 1-19.
- Thorpe, G. L., & Favia, A. (2012). Data analysis using Item Response Theory Methodology: An introduction to selected programs and applications. *Psychology Faculty Scholarship*. Paper 20. [https://digitalcommons.library.umaine.edu/psy\\_facpub/20](https://digitalcommons.library.umaine.edu/psy_facpub/20)
- Toland, M. D. (2014). Practical guide to conducting an Item Response Theory analysis. *Journal of Early Adolescence, 34*(1), 120-151. doi: 10.1177/0272431613511332
- Zhu, W. (1996). Should total scores from a rating scale be used directly? *Research Quarterly for Exercise and Sport, 67*(3), 363-372.



## ΠΑΡΑΡΤΗΜΑ Α

Y-CATS Conflict 1PL model

IRTPRO Version 4.1

Output generated by IRTPRO estimation engine Version 5.20 (32-bit)

Project:	Y-CATS Conflict 1PL model
Description:	
Date:	30 April 2019
Time:	07:20 PM

### Table of Contents

2PL Model Item Parameter Estimates for Group 1, logit:  $a\theta + c$  or  $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and  $\chi^2$ 's for Group 1

Group Parameter Estimates

Marginal fit ( $\chi^2$ ) and Standardized LD  $\chi^2$  Statistics for Group 1

Item Information Function Values for Group 1 at 15 Values of  $\theta$  from -2.8 to 2.8

Likelihood-based Values and Goodness of Fit Statistics

Summary of the Data and Control Parameters

### 2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label	a	s.e.	c	s.e.	b	s.e.		
1	y2co	0	1.21	0.05	1	-1.64	0.10	1.35	0.09
2	y5co	0	1.21	0.05	2	-0.52	0.08	0.43	0.07
3	y8co	0	1.21	0.05	3	-0.41	0.08	0.34	0.07
4	y11co	0	1.21	0.05	4	-0.54	0.08	0.44	0.07
5	y14co	0	1.21	0.05	5	-1.14	0.09	0.94	0.08
6	y17co	0	1.21	0.05	6	-0.93	0.09	0.77	0.08
7	y20co	0	1.21	0.05	7	-0.64	0.08	0.53	0.07
8	y22co	0	1.21	0.05	9	1.22	0.09	-1.01	0.09
9	y24co	0	1.21	0.05	10	-0.86	0.09	0.71	0.08
10	y26co	0	1.21	0.05	11	-0.46	0.08	0.38	0.07

### Summed-Score Based Item Diagnostic Tables and $\chi^2$ 's for Group 1 (Back to TOC)

#### S- $\chi^2$ Item Level Diagnostic Statistics

Item	Label	$\chi^2$	df	Probability
1	y2co	11.83	8	0.1584
2	y5co	9.13	8	0.3331
3	y8co	11.93	8	0.1541
4	y11co	9.95	8	0.2702
5	y14co	18.23	8	0.0195
6	y17co	27.99	8	0.0005
7	y20co	18.30	8	0.0191
8	y22co	34.75	7	0.0001
9	y24co	12.93	8	0.1138
10	y26co	10.81	8	0.2122

#### Group Parameter Estimates (Back to TOC)

Group	Label	$\mu$	s.e.	$\sigma^2$	s.e.	$\sigma$	s.e.
1	Group 1	0.00	----	1.00	----	1.00	----

#### Marginal fit ( $\chi^2$ ) and Standardized LD $\chi^2$ Statistics for Group 1 (Back to TOC)

Item	Label	Marginal																	
		$\chi^2$	1	2	3	4	5	6	7	8	9								
1	y2co	0.0																	
2	y5co	0.0	0.2																
3	y8co	0.0	2.5	-0.7															
4	y11co	0.0	0.4	-0.6	-0.5														
5	y14co	0.0	14.9	-0.6	-0.2	-0.3													
6	y17co	0.0	4.4	4.3	-0.6	2.1	-0.7												
7	y20co	0.0	3.5	-0.7	0.5	1.2	1.1	1.6											
8	y22co	0.0	-0.1	2.3	15.2	1.6	2.0	10.5	-0.7										
9	y24co	0.0	3.2	-0.1	0.4	-0.7	17.2	-0.4	7.9	1.4									
10	y26co	0.0	1.2	-0.6	3.5	0.4	-0.1	1.4	0.7	0.6	0.5								

#### Item Information Function Values for Group 1 at 15 Values of $\theta$ from -2.8 to 2.8 (Back to TOC)

Item	Label	$\theta$ :														
		-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	y2co	0.01	0.02	0.02	0.04	0.06	0.09	0.14	0.20	0.27	0.33	0.36	0.36	0.31	0.25	0.18
2	y5co	0.03	0.04	0.07	0.11	0.16	0.22	0.29	0.34	0.37	0.35	0.30	0.23	0.17	0.11	0.07
3	y8co	0.03	0.05	0.08	0.12	0.17	0.24	0.30	0.35	0.37	0.34	0.28	0.21	0.15	0.10	0.07
4	y11co	0.03	0.04	0.07	0.11	0.16	0.22	0.28	0.34	0.37	0.35	0.30	0.23	0.17	0.11	0.08
5	y14co	0.02	0.02	0.04	0.06	0.10	0.14	0.20	0.27	0.33	0.36	0.36	0.31	0.25	0.18	0.13
6	y17co	0.02	0.03	0.05	0.07	0.11	0.17	0.23	0.30	0.35	0.37	0.34	0.29	0.22	0.16	0.11
7	y20co	0.03	0.04	0.06	0.10	0.14	0.20	0.27	0.33	0.36	0.36	0.31	0.25	0.18	0.13	0.08
8	y22co	0.13	0.19	0.26	0.32	0.36	0.36	0.32	0.26	0.19	0.13	0.09	0.06	0.04	0.02	0.01
9	y24co	0.02	0.03	0.05	0.08	0.12	0.17	0.24	0.31	0.35	0.36	0.34	0.28	0.21	0.15	0.10
10	y26co	0.03	0.05	0.07	0.11	0.16	0.23	0.29	0.35	0.37	0.34	0.29	0.22	0.16	0.11	0.07
Test Information:		1.34	1.52	1.77	2.11	2.54	3.04	3.67	4.04	4.31	4.29	3.96	3.44	2.86	2.33	1.90
Expected s.e.:		0.86	0.81	0.75	0.69	0.63	0.57	0.53	0.50	0.48	0.48	0.50	0.54	0.59	0.66	0.72

Y-CATS Conflict 1PL model

**Marginal Reliability for Response Pattern Scores: 0.70****Likelihood-based Values and Goodness of Fit Statistics** (Back to TOC)

Statistics based on the loglikelihood

-2loglikelihood:	10516.15
Akaike Information Criterion (AIC):	10538.15
Bayesian Information Criterion (BIC):	10590.97

Statistics based on the full item x item x ... classification

The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one- and two-way marginal tables

$M_2$	Degrees of freedom	Probability	RMSEA
150.05	44	0.0001	0.05

Note:  $M_2$  is based on full marginal tables.

Note: Model-based weight matrix is used.

**Summary of the Data and Control Parameters** (Back to TOC)

Sample Size	900
Number of Items	10
Number of Dimensions	1

Y-CATS Conflict 2PL model

**IRTPRO Version 4.1**  
**Output generated by IRTPRO estimation engine Version 5.20 (32-bit)**

Project:	Y-CATS Conflict 2PL model
Description:	
Date:	30 April 2019
Time:	07:20 PM

**Table of Contents**

2PL Model Item Parameter Estimates for Group 1, logit:  $a\theta + c$  or  $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and  $\chi^2$ 's for Group 1

Group Parameter Estimates

Marginal fit ( $\chi^2$ ) and Standardized LD  $\chi^2$  Statistics for Group 1

Item Information Function Values for Group 1 at 15 Values of  $\theta$  from -2.8 to 2.8

Likelihood-based Values and Goodness of Fit Statistics

Summary of the Data and Control Parameters

**2PL Model Item Parameter Estimates for Group 1, logit:  $a\theta + c$  or  $a(\theta - b)$ . (Back to TOC)**

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	y2co	2	1.70	0.19	1	-1.90	0.15
2	y5co	4	0.98	0.11	3	-0.49	0.08
3	y8co	6	1.15	0.12	5	-0.41	0.08
4	y11co	8	1.19	0.13	7	-0.54	0.09
5	y14co	10	1.79	0.19	9	-1.36	0.13
6	y17co	12	0.78	0.10	11	-0.82	0.08
7	y20co	14	1.61	0.17	13	-0.74	0.10
8	y22co	16	0.66	0.10	15	1.04	0.08
9	y24co	18	1.69	0.18	17	-1.00	0.11
10	y26co	20	1.23	0.13	19	-0.47	0.09

**Summed-Score Based Item Diagnostic Tables and  $\chi^2$ 's for Group 1 (Back to TOC)**

**S- $\chi^2$  Item Level Diagnostic Statistics**

Item	Label	$\chi^2$	d.f.	Probability
1	y2co	6.43	8	0.6007
2	y5co	5.08	8	0.7501
3	y8co	10.84	8	0.2106
4	y11co	10.09	8	0.2584
5	y14co	8.94	8	0.3491
6	y17co	11.31	8	0.1845
7	y20co	12.09	8	0.1471
8	y22co	8.10	8	0.4251
9	y24co	6.16	8	0.6301
10	y26co	11.85	8	0.1574

**Group Parameter Estimates (Back to TOC)**

Group	Label	$\mu$	s.e.	$\sigma^2$	s.e.	$\sigma$	s.e.
1	Group 1	0.00	-----	1.00	-----	1.00	-----

**Marginal fit ( $\chi^2$ ) and Standardized LD  $\chi^2$  Statistics for Group 1 (Back to TOC)**

Item	Label	Marginal									
		$\chi^2$	1	2	3	4	5	6	7	8	9
1	y2co	0.0									
2	y5co	0.0	0.9								
3	y8co	0.0	-0.0	0.4							
4	y11co	0.0	-0.7	-0.4	-0.6						
5	y14co	0.0	0.8	-0.7	-0.4	-0.2					
6	y17co	0.0	2.2	-0.7	2.8	-0.7	-0.5				
7	y20co	0.0	-0.6	-0.7	-0.6	-0.5	1.3	-0.2			
8	y22co	0.0	-0.6	-0.3	2.5	-0.5	-0.4	-0.4	1.0		
9	y24co	0.0	-0.4	0.4	-0.7	0.9	1.2	-0.7	-0.5	-0.7	
10	y26co	0.0	-0.7	0.4	3.1	0.3	-0.2	-0.7	-0.7	-0.1	-0.7

**Item Information Function Values for Group 1 at 15 Values of  $\theta$  from -2.8 to 2.8 (Back to TOC)**

Item	Label	$\theta$ :														
		-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	y2co	0.00	0.01	0.01	0.03	0.05	0.10	0.19	0.33	0.51	0.67	0.72	0.61	0.43	0.26	0.15
2	y5co	0.04	0.05	0.07	0.10	0.13	0.16	0.20	0.22	0.24	0.23	0.21	0.18	0.15	0.11	0.08
3	y8co	0.03	0.05	0.08	0.11	0.16	0.22	0.28	0.32	0.33	0.31	0.26	0.21	0.15	0.10	0.07
4	y11co	0.03	0.04	0.07	0.10	0.15	0.21	0.28	0.33	0.36	0.34	0.29	0.23	0.17	0.12	0.08
5	y14co	0.01	0.01	0.02	0.05	0.09	0.17	0.32	0.52	0.73	0.80	0.69	0.48	0.28	0.15	0.08
6	y17co	0.03	0.04	0.05	0.06	0.08	0.09	0.11	0.13	0.14	0.15	0.15	0.15	0.13	0.12	0.10
7	y20co	0.01	0.03	0.05	0.09	0.16	0.27	0.42	0.57	0.65	0.60	0.46	0.31	0.18	0.10	0.06
8	y22co	0.09	0.10	0.11	0.11	0.11	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.03	0.02
9	y24co	0.01	0.02	0.03	0.07	0.13	0.23	0.38	0.56	0.69	0.69	0.55	0.37	0.22	0.12	0.07
10	y26co	0.03	0.05	0.07	0.11	0.17	0.23	0.30	0.36	0.38	0.36	0.30	0.23	0.16	0.11	0.07
Test Information:		1.28	1.39	1.56	1.82	2.22	2.79	3.56	4.42	5.10	5.23	4.70	3.81	2.91	2.23	1.77
Expected s.e.:		0.89	0.85	0.80	0.74	0.67	0.60	0.53	0.48	0.44	0.44	0.46	0.51	0.59	0.67	0.75

1/2

## Y-CATS Conflict 2PL model

**Marginal Reliability for Response Pattern Scores: 0.71**

**Likelihood-based Values and Goodness of Fit Statistics** (Back to TOC)

Statistics based on the loglikelihood

-2loglikelihood:	10439.13
Akaike Information Criterion (AIC):	10479.13
Bayesian Information Criterion (BIC):	10575.18

Statistics based on the full item x item x ... classification

The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one- and two-way marginal tables

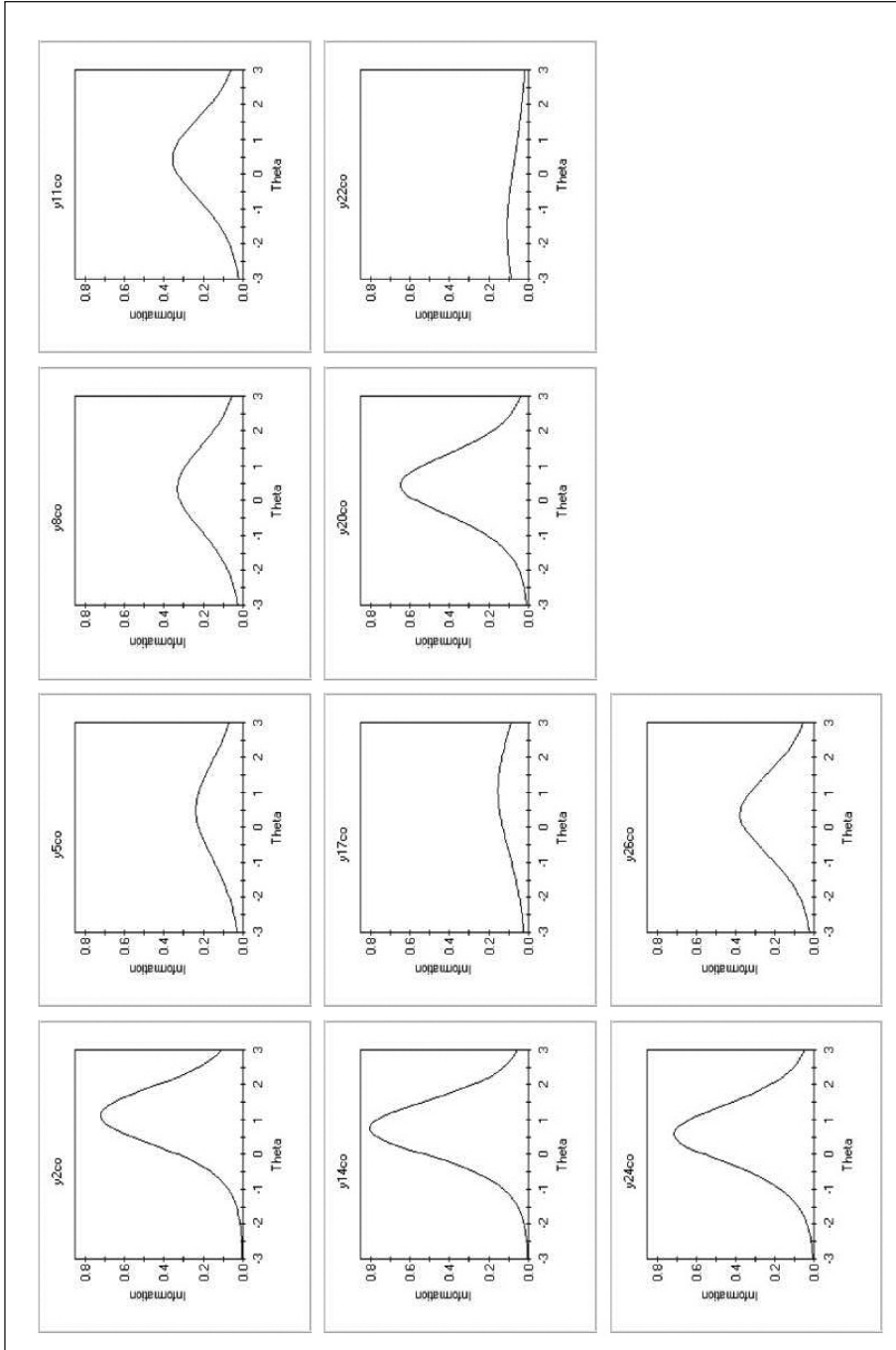
$M_2$	Degrees of freedom	Probability	RMSEA
69.88	35	0.0004	0.03

Note:  $M_2$  is based on full marginal tables.

Note: Model-based weight matrix is used.

**Summary of the Data and Control Parameters** (Back to TOC)

Sample Size	900
Number of Items	10
Number of Dimensions	1



## ΠΑΡΑΡΤΗΜΑ Β

### Θεωρία Απόκρισης Στοιχείου με το πρόγραμμα ltm

#### 1. ΤΟ ΠΕΡΙΒΑΛΛΟΝ R ΚΑΙ RStudio

Το πρόγραμμα R είναι ένα σύνθετο περιβάλλον προγραμματισμού και εκτέλεσης στατιστικών αναλύσεων. Ανάμεσα στα πολλά πλεονέκτημα του είναι ότι διανέμεται δωρεάν, εξελίσσεται αρκετά γρήγορα, ενημερώνεται σε τακτική βάση και διαθέτει εφαρμογές (libraries) για κάθε είδους ανάλυση. Το πρόγραμμα μπορεί να μεταφορτωθεί από τη διεύθυνση <https://cran.r-project.org/> και μπορεί να εγκατασταθεί σε διάφορες πλατφόρμες (Windows, OS, Linux). Το RStudio αποτελεί ένα κέλυφος μέσω του οποίου μπορεί να πραγματοποιηθούν οι λειτουργίες του R σε γραφικό περιβάλλον. Το RStudio μπορεί να μεταφορτωθεί και να εγκατασταθεί δωρεάν από τη διεύθυνση <https://www.rstudio.com/products/RStudio/>.

#### 2. ΕΙΣΑΓΩΓΗ ΚΑΙ ΕΞΕΤΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Αρχικά θα πρέπει να γίνει εισαγωγή των δεδομένων προς ανάλυση. Για το λόγο αυτό θα χρησιμοποιηθεί το πρόγραμμα “haven” [install.packages(“haven”)] το οποίο επιτρέπει την εισαγωγή δεδομένων στο περιβάλλον R από διάφορες εμπορικές στατιστικές εφαρμογές όπως SPSS, SAS και STATA.

```
library(haven)
## Warning: package 'haven' was built under R version 3.5.3
```

Στο επόμενο βήμα καθορίζεται η διαδρομή του φάκελου στον οποίο υπάρχει το αρχείο των δεδομένων (“YCATS conflict.sav”) και στη συνέχεια εισάγουμε τα δεδομένα με την εντολή “read\_spss” και εκχωρούμε το αποτέλεσμα στο αντικείμενο “conflict”.

```
setwd(“C:/Users/079/Desktop”)
conflict<-read_spss(“YCATS conflict.sav”)
```

Με τις παρακάτω τρεις εντολές ελέγχουμε εάν τα δεδομένα έχουν εισαχθεί ορθά. Πιο συγκεκριμένα, με την πρώτη εντολή εμφανίζονται τα ονόματα των μεταβλητών, με τη δεύτερη ο αριθμός των περιπτώσεων και ο αριθμός των μεταβλητών, ενώ η τρίτη εμφανίζει τις πρώτες έξι καταγραφές.

```

names(conflict)
## [1] "y2co" "y5co" "y8co" "y11co" "y14co" "y17co" "y20co" "y22co"
## [9] "y24co" "y26co"
dim(conflict)
## [1] 900 10
head(conflict)
## # A tibble: 6 x 10
##   y2co   y5co   y8co  y11co  y14co  y17co  y20co  y22co  y24co
##   <dbl+1> <dbl+1> <dbl+1> <dbl+1> <dbl+1> <dbl+1> <dbl+1> <dbl+1> <dbl+1>
##   1 0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  1 [yes]  0 [no]  0 [no]  0 [no]
##   2 0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]
##   3 0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  0 [no]  1 [yes]  0 [no]
##   4 0 [yes] 0 [no]  1 [yes] 0 [no]  1 [yes] 1 [yes] 0 [no]  0 [no]  1 [yes]
##   5 0 [no]  0 [no]  0 [no]  0 [no]  1 [yes] 1 [yes] 0 [no]  0 [no]  0 [no]
##   6 0 [no]  1 [yes] 0 [no]  1 [yes] 0 [no]  0 [no]  0 [no]  0 [no]  0 [no]
## # ... with 1 more variable: y26co <dbl+1>

```

Από τα αποτελέσματα φαίνεται ότι έχουν εισαχθεί 10 μεταβλητές (y2co έως και y26co) για 900 περιπτώσεις.

### 3. ΕΚΤΕΛΕΣΗ ΑΝΑΛΥΣΕΩΝ ΘΕΩΡΙΑΣ ΑΠΟΚΡΙΣΗΣ ΣΤΟΙΧΕΙΟΥ

Στη συνέχεια θα πρέπει να εγκατασταθεί [install.packages("ltm")] και να ενεργοποιηθεί η βιβλιοθήκη "ltm" με την οποία θα πραγματοποιηθούν οι αναλύσεις της ΘΑΣ.

```

library(ltm)
## Warning: package 'ltm' was built under R version 3.5.3
## Loading required package: MASS
## Warning: package 'MASS' was built under R version 3.5.3
## Loading required package: msm
## Warning: package 'msm' was built under R version 3.5.3
## Loading required package: polycor

```

Παρατηρούμε ότι έχει ενεργοποιηθεί η βιβλιοθήκη "ltm" μαζί με τις συνοδευτικές βιβλιοθήκες. Η βιβλιοθήκη "ltm" έχει δημιουργηθεί και επικαιροποιηθεί πρόσφατα από τον Dimitri Rizopoulos (2018) και το συνοδευτικό της εγχειρίδιο βρίσκεται στη διεύθυνση <https://cran.r-project.org/web/packages/ltm/ltm.pdf>.

Κατά την εφαρμογή της ΘΑΣ για τη μελέτη των δέκα δηλώσεων της αντιλαμβανό-

μενης συγκρουσιακής σχέσης των μαθητών με τον εκπαιδευτικό θα εξεταστούν δύο μοντέλα, μίας και δύο παραμέτρων. Στο μοντέλο μιας παραμέτρου θα εκχωρηθεί το όνομα “conflict\_1PL” και στο δεύτερο το “conflict\_2PL”. Η εντολή “rasch” χρησιμοποιείται για τον υπολογισμό του μοντέλου μίας παραμέτρου. Το μόνο που χρειάζεται είναι να δηλώσουμε το όνομα του αρχείου που περιέχει τα δεδομένα. Με την επόμενη εντολή “summary” εμφανίζονται οι δείκτες προσαρμογής του μοντέλου καθώς και οι παράμετροι του. Αξίζει να σημειωθεί ότι εάν χρειαζόνταν να εξεταστεί ένα Rasch μοντέλο στο οποίο η τιμή της διάκρισης να είναι μονάδα για όλα τα στοιχεία η εντολή θα συντάσσονταν ως εξής: `conflict_Rasch<-rasch(conflict, constraint = cbind(ncol(conflict)+1,1))`.

```

conflict_1PL<-rasch(conflict)
summary(conflict_1PL)
##
## Call:
## rasch(data = conflict)
##
## Model Summary:
##   log.Lik   AIC   BIC
## -5258.076 10538.15 10590.98
##
## Coefficients:
##           value   std.err   z.vals
## Dffclt.y2co  1.3528   0.0919  14.7253
## Dffclt.y5co  0.4326   0.0733   5.8994
## Dffclt.y8co  0.3400   0.0725   4.6910
## Dffclt.y11co 0.4422   0.0734   6.0221
## Dffclt.y14co 0.9387   0.0814  11.5387
## Dffclt.y17co 0.7689   0.0780   9.8514
## Dffclt.y20co 0.5323   0.0745   7.1480
## Dffclt.y22co -1.0064   0.0827 -12.1729
## Dffclt.y24co 0.7113   0.0771   9.2297
## Dffclt.y26co 0.3838   0.0729   5.2672
## Dscrmn      1.2092   0.0474  25.5220
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.087
## quasi-Newton: BFGS

```



Αρχικά εμφανίζονται οι δείκτες προσαρμογής του μοντέλου μίας παραμέτρου. Η τιμή  $\log.Lik$  (δηλαδή  $\log$  likelihood) εάν πολλαπλασιαστεί με το -2 τότε δίνει την ίδια ακριβώς τιμή με αυτή του IRTPRO ( $-2 \cdot -5258.076 = 10516.15$ ). Επίσης δίνονται οι τιμές των δεικτών AIC και BIC. Οι δείκτες αυτοί μπορεί να χρησιμοποιηθούν για τη σχετική προσαρμογή του μοντέλου σε σχέση με το μοντέλο των δύο παραμέτρων. Στη συνέχεια παρουσιάζονται οι 11 παράμετροι του μοντέλου (10 τιμές δυσκολίας, μία για κάθε δήλωση και μια τιμή διάκρισης κοινή για όλα τα στοιχεία), μαζί με τα τυπικά τους σφάλματα και τις z-τιμές. Τιμές z μεγαλύτερες της απόλυτης τιμής 1.961 θεωρούνται στατιστικά σημαντικές σε επίπεδο σημαντικότητας .05.

Το δεύτερο υποψήφιο μοντέλο προς εξέταση είναι το μοντέλο των δύο παραμέτρων. Η εντολή που χρησιμοποιείται είναι "ltm". Μια διαφορά από την εντολή "rasch" είναι ότι το όνομα του αρχείου δεδομένων συνοδεύεται από το "~ z1", το οποίο δηλώνει ότι κάθε στοιχείο θα έχει τη δική του τιμή διάκρισης και δυσκολίας.

```

conflict_2PL<-ltm(conflict ~ z1, IRT.param = TRUE)
summary(conflict_2PL)
##
## Call:
## ltm(formula = conflict ~ z1, IRT.param = TRUE)
##
## Model Summary:
## log.Lik   AIC   BIC
## -5219.56 10479.12 10575.17
##
## Coefficients:
##           value      std.err  z.vals
## Dffclt.y2co  1.1189    0.0922   12.1416
## Dffclt.y5co  0.5046    0.0933    5.4107
## Dffclt.y8co  0.3552    0.0780    4.5529
## Dffclt.y11co 0.4510    0.0791    5.7046
## Dffclt.y14co 0.7606    0.0723   10.5161
## Dffclt.y17co 1.0512    0.1508    6.9698
## Dffclt.y20co 0.4575    0.0663    6.9004
## Dffclt.y22co -1.5804    0.2462   -6.4194
## Dffclt.y24co 0.5954    0.0686    8.6738
## Dffclt.y26co 0.3844    0.0756    5.0875
## Dscrmn.y2co  1.6991    0.1884    9.0181
## Dscrmn.y5co  0.9760    0.1129    8.6413
## Dscrmn.y8co  1.1546    0.1252    9.2235
## Dscrmn.y11co 1.1944    0.1280    9.3294

```

```

## Dscrmn.y14co  1.7950    0.1907    9.4104
## Dscrmn.y17co  0.7847    0.1047    7.4953
## Dscrmn.y20co  1.6106    0.1653    9.7423
## Dscrmn.y22co  0.6569    0.1051    6.2521
## Dscrmn.y24co  1.6879    0.1751    9.6410
## Dscrmn.y26co  1.2320    0.1309    9.4129
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.00034
## quasi-Newton: BFGS

```

Υπάρχει επίσης η δυνατότητα να ζητηθούν μόνο οι παράμετροι του μοντέλου με την εντολή “coef”

```

coef(conflict_2PL)
##          Dffclt          Dscrmn
## y2co      1.1189195    1.6990933
## y5co      0.5045909    0.9759715
## y8co      0.3552037    1.1545754
## y11co     0.4510080    1.1943535
## y14co     0.7605825    1.7949620
## y17co     1.0512185    0.7846877
## y20co     0.4575494    1.6105763
## y22co    -1.5803977    0.6568986
## y24co     0.5953682    1.6878987
## y26co     0.3843781    1.2319758

```

Η σύγκριση των δύο μοντέλων μπορεί να πραγματοποιηθεί με την εντολή “anova”. Παρά το γεγονός ότι το όνομα της εντολής αυτής παραπέμπει σε ανάλυση διακύμανσης στην πραγματικότητα συγκρίνει δύο εμφωλευμένα μοντέλα (nested models). Ενημερωτικά, η εντολή για την εκτέλεση ανάλυση διακύμανσης είναι “aov”.

```

anova(conflict_1PL, conflict_2PL)
## Likelihood Ratio Table
##          AIC          BIC          log.Lik  LRT  df  p.value
## conflict_1PL 10538.15 10590.98 -5258.08
## conflict_2PL 10479.12 10575.17 -5219.56 77.03  9  <0.001

```

Από τα αποτελέσματα της σύγκρισης προκύπτει ότι η τιμή του Likelihood Ratio Test (LRT) είναι 77.03, η οποία ακολουθεί τη  $\chi^2$  κατανομή με 9 βαθμούς ελευθερίας, είναι στατιστικά σημαντική. Συνεπώς το μοντέλο των δύο παραμέτρων έχει καλύτερη προσαρμογή σε σχέση με το μοντέλο της μίας παραμέτρου.

Παράλληλα με την επιλογή του μοντέλου με την καλύτερη προσαρμογή θα πρέπει να εξεταστεί και η προσαρμογή του κάθε ενός από τα δέκα στοιχεία. Ο τρόπος υπολογισμού της προσαρμογής διαφέρει από αυτή του IRTPRO. Πιο συγκεκριμένα, στηρίζεται στη δοκιμασία  $\chi^2$ . Το στατιστικό αυτό υπολογίζεται συγκρίνοντας τη συχνότητα των συμμετεχόντων που ορθά αποκρίθηκαν στο στοιχείο για μικρά τμήματα του λανθάνοντος χαρακτηριστικού (10 είναι η προεπιλογή στο ltm) με το αναμενόμενο ποσοστό που προβλέπεται από το μοντέλο. Επίσης εάν στην εντολή “item.fit” προστεθεί το “simulate.p.value = TRUE” τότε δεν χρησιμοποιείται η θεωρητική  $\chi^2$  κατανομή, αλλά η εμπειρική από την αναδειγματοληψία (bootstrapping) εκατό δειγμάτων. Ο αριθμός των τμημάτων στα οποία θα χωριστεί η λανθάνουσα μεταβλητή (“G”) καθώς και ο αριθμός του μεγέθους της αναδειγματοληψίας (“B”) μπορεί να οριστεί από τον ερευνητή.

```

item.fit(conflict_2PL, G = 10, simulate.p.value = TRUE, B = 100)
## Item-Fit Statistics and P-values
##
## Call:
## ltm(formula = conflict ~ z1, IRT.param = TRUE)
##
## Alternative: Items do not fit the model
## Ability Categories: 10
## Monte Carlo samples: 100
##
##           X^2 Pr(>X^2)
## y2co    34.9493  0.3267
## y5co    29.7811  0.3267
## y8co    25.2071  0.7129
## y11co   25.8459  0.7228
## y14co   34.4222  0.5743
## y17co   21.9355  0.3663
## y20co   49.8994  0.0792
## y22co   10.1073  0.9604
## y24co   33.2436  0.703
## y26co   39.4806  0.1584

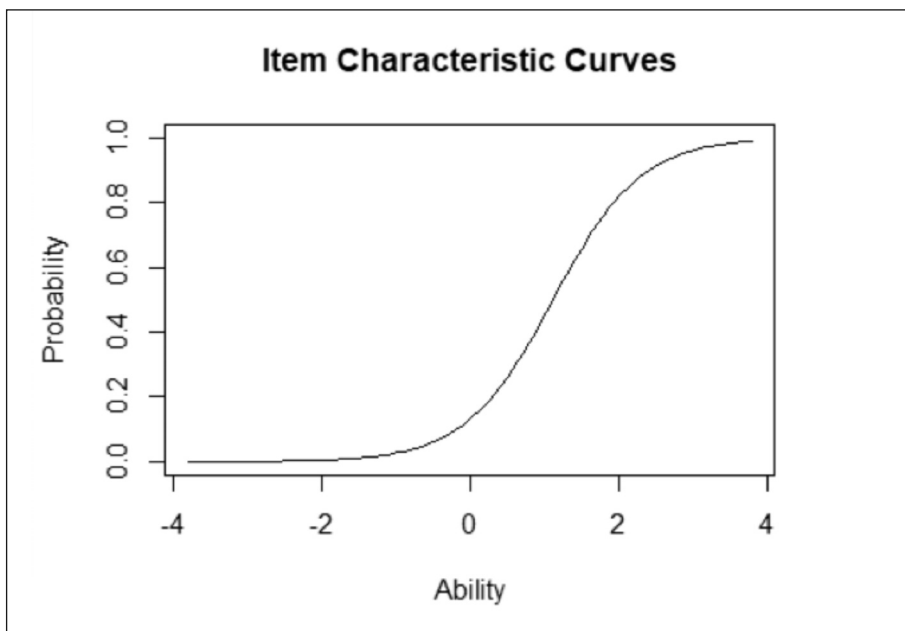
```

Από τα αποτελέσματα φαίνεται ότι και τα δέκα στοιχεία της συγκρουσιακής σχέσης έχουν καλή προσαρμογή καθώς καμία  $p$ -τιμή δεν είναι μικρότερη από το .05.

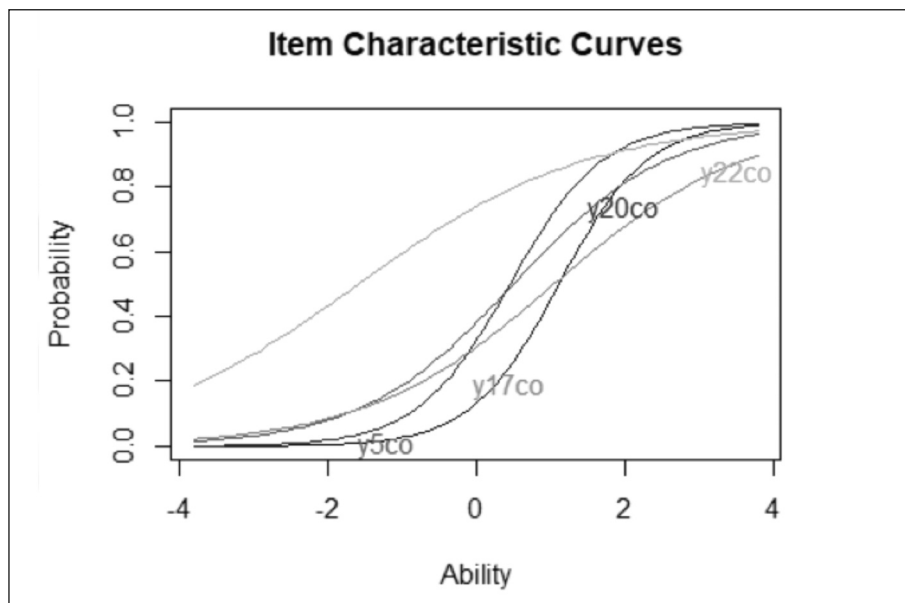
#### 4. ΔΗΜΙΟΥΡΓΙΑ ΚΑΤΑΛΛΗΛΩΝ ΓΡΑΦΗΜΑΤΩΝ

Για την καλύτερη κατανόηση των αποτελεσμάτων ο ερευνητής μπορεί εύκολα να δημιουργήσει διάφορα γραφήματα. Οι παρακάτω εντολές αφορούν τη δημιουργία χαρακτηριστικής καμπύλης στοιχείου (item characteristic curve, ICC), για ένα στοιχείο, για μερικά στοιχεία και για όλα τα στοιχεία της μελέτης.

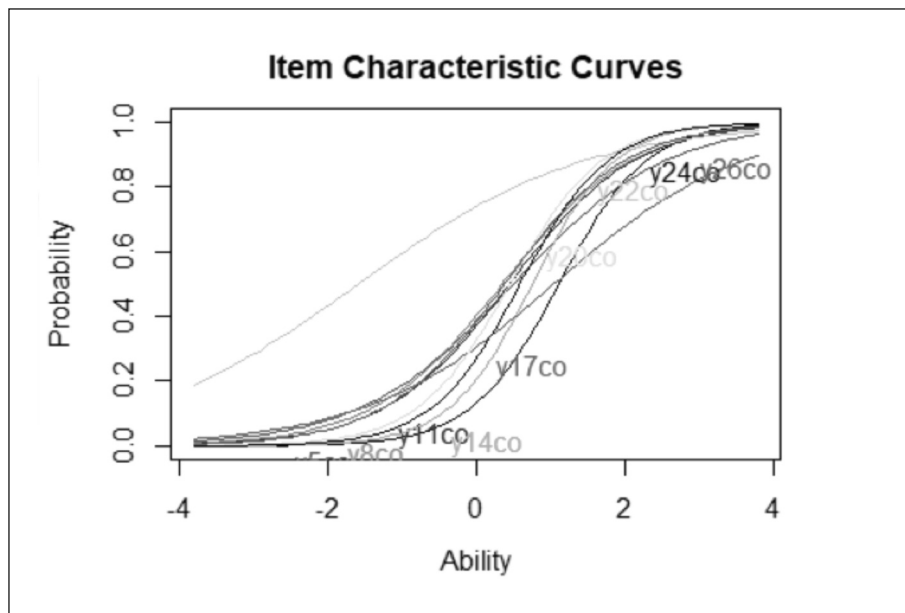
```
plot(conflict_2PL, type="ICC", items=1)
```



```
plot(conflict_2PL, type="ICC", items=c(1, 2, 6:8))
```

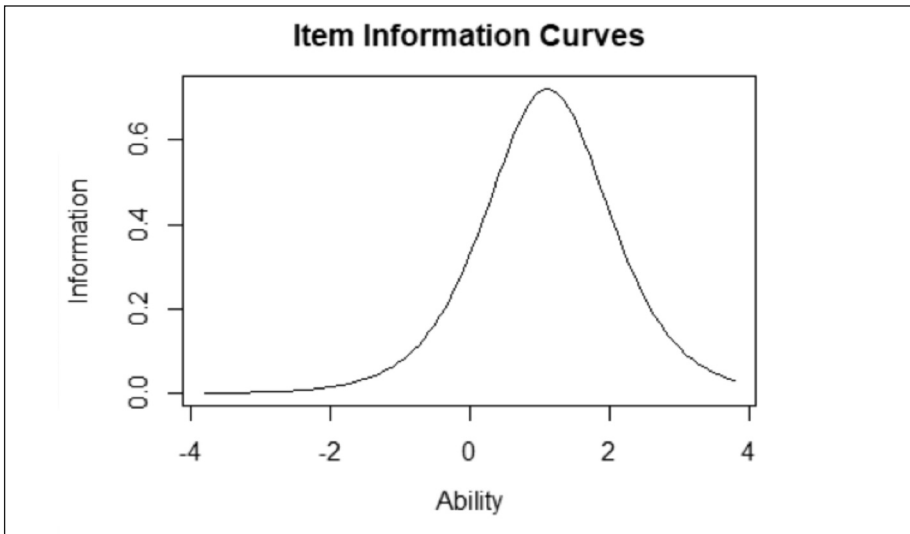


```
plot(conflict_2PL, type="ICC")
```

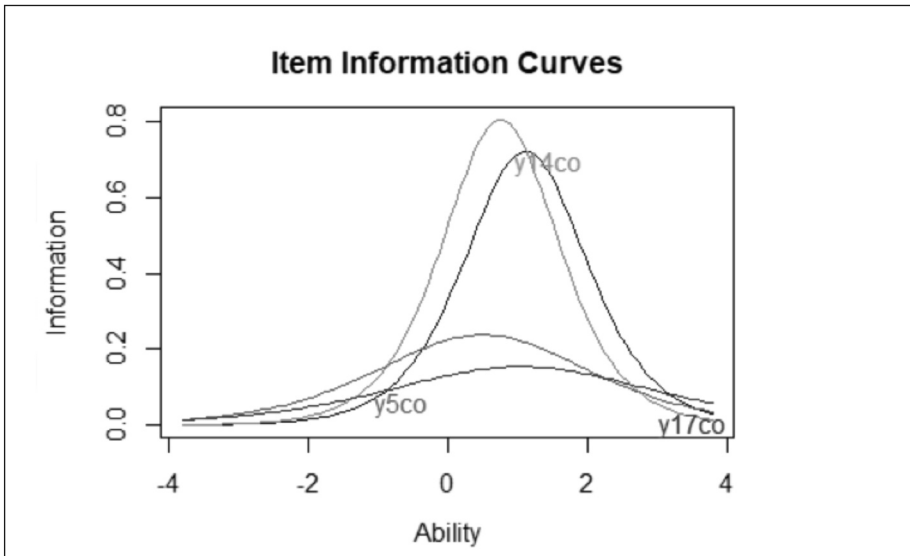


Οι παρακάτω εντολές αφορούν τη δημιουργία καμπύλης πληροφορίας στοιχείου (item information curve, IIC), για ένα στοιχείο, για μερικά στοιχεία, για όλα τα στοιχεία της μελέτης, καθώς και τη συνολική πληροφορία του λανθάνοντος χαρακτηριστικού.

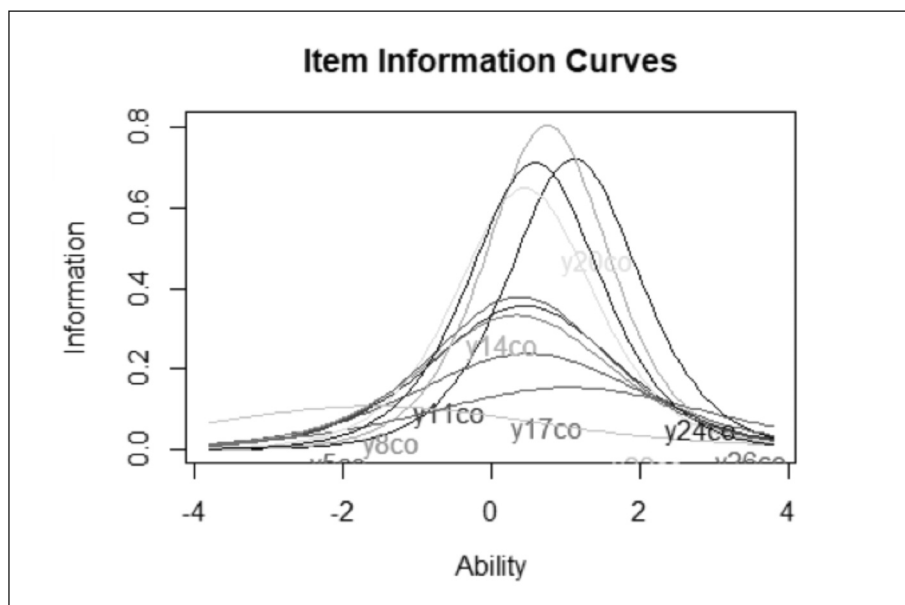
```
plot(conflict_2PL, type="IIC", items=1)
```



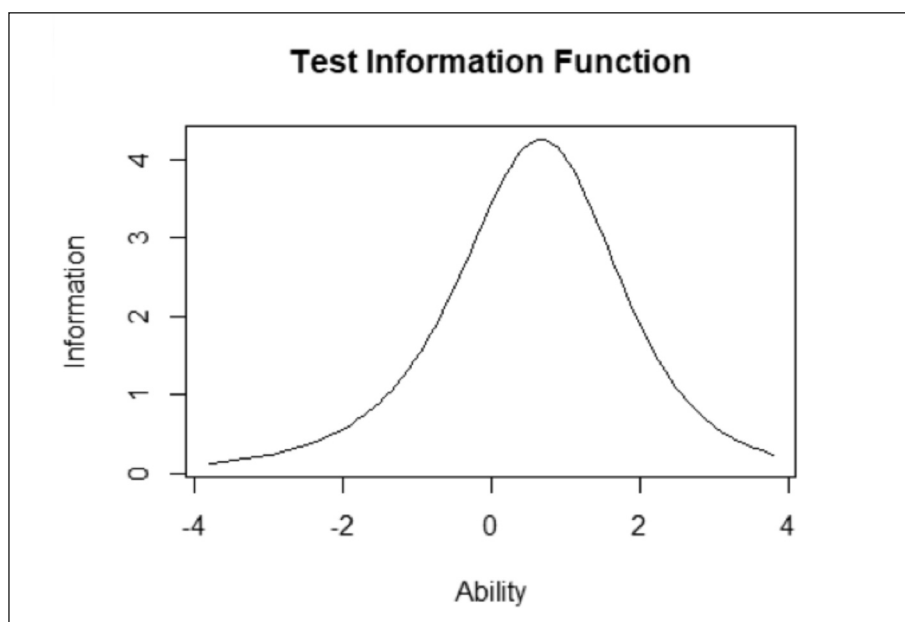
```
plot(conflict_2PL, type="IIC", items=c(1, 2, 5:6))
```



```
plot(conflict_2PL, type="IIC")
```



```
plot(conflict_2PL, type="IIC", items=0)
```



## 5. ΤΙΜΕΣ ΛΑΝΘΑΝΟΝΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ

Οι τιμές του λανθάνοντος χαρακτηριστικού “συγκρουσιακή σχέση” μπορεί να υπολογιστούν με την εντολή “factor scores”. Στο συγκεκριμένο παράδειγμα η μέθοδος υπολογισμού είναι η “EAP” (Expected a Priori) και το αποτέλεσμα του υπολογισμού θα εκχωρηθεί στο αντικείμενο “conflict\_lv”. Με τη δεύτερη εντολή ζητείται να εμφανιστούν τα δέκα πρώτα μοτίβα απόκρισης και οι τιμές του λανθάνοντος χαρακτηριστικού για κάθε μοτίβο.

```

conflict_lv <- factor.scores.ltm(conflict_2PL, method="EAP")
conflict_lv$score.dat[1:10, ]
##      y2co y5co y8co y11co y14co y17co y20co y22co y24co y26co Obs      Exp
## 1      0      0      0      0      0      0      0      0      0      0  52 53.111116
## 2      0      0      0      0      0      0      0      0      0      0   1  4 8.286440
## 3      0      0      0      0      0      0      0      0      0      1   0  6 3.389690
## 4      0      0      0      0      0      0      0      0      0      1   1  3 1.111432
## 5      0      0      0      0      0      0      0      0      1      0   0 79 66.278741
## 6      0      0      0      0      0      0      0      0      1      0   1 16 14.199599
## 7      0      0      0      0      0      0      0      0      1      1   0  7  6.418698
## 8      0      0      0      0      0      0      0      0      1      1   1  4  2.663320
## 9      0      0      0      0      0      0      0      1      0      0   0  4  4.688418
## 10     0      0      0      0      0      0      1      1      0      0   8  8.733692

##              z1          se.z1
## 1 -1.3918699  0.6881938
## 2 -0.8768157  0.6072415
## 3 -0.7156486  0.5822261
## 4 -0.3384442  0.5267714
## 5 -1.1012282  0.6426875
## 6 -0.6487169  0.5720202
## 7 -0.5051556  0.5505807
## 8 -0.1641268  0.5043250
## 9 -0.7420424  0.5862828
## 10 -0.5287429  0.5540556

```

Εξέταση του δεύτερου (##2) και τρίτου μοτίβου (##3) αποκαλύπτει το διαφορετικό τρόπο βαθμολόγησης σε σχέση με την Κλασική Θεωρία Μέτρησης. Πιο αναλυτικά, το δεύτερο μοτίβο απόκρισης δείχνει μια θετική μόνο απόκριση στο στοιχείο “y26co”, η οποία παρατηρήθηκε σε τέσσερα άτομα (στήλη “Obs”). Το τρίτο μοτίβο απόκρισης δείχνει επίσης μια θετική απόκριση αλλά στο στοιχείο “y24co” από έξι



άτομα (στήλη “Obs”). Εάν χρησιμοποιούνταν η Κλασική Θεωρία Μέτρησης τα άτομα αυτά θα λάμβαναν την ίδια ακριβώς βαθμολογία. Στο πλαίσιο όμως της ΘΑΣ επειδή τα άτομα αποκρίθηκαν θετικά σε δύο ξεχωριστά στοιχεία τα οποία έχουν διαφορετικές τιμές διάκρισης και δυσκολίας λαμβάνουν διαφορετικές τιμές. Τα τέσσερα άτομα του δεύτερου μοτίβου λαμβάνουν την τιμή -0.8768 (στήλη z1) και τα έξι άτομα του τρίτου την τιμή -0.7156. Η διαφοροποίηση αυτή στον τρόπο βαθμολόγησης αποτελεί ένα σημαντικό πλεονέκτημα επιλογής της ΘΑΣ.

Με την παρακάτω εντολή υπολογίζονται οι λανθάνουσες τιμές για κάθε συμμετέχοντα και τα αποτελέσματα εκχωρούνται στο αντικείμενο “conflict\_latent\_scores”. Στη συνέχεια δίνονται τα περιγραφικά στατιστικά του λανθάνοντος χαρακτηριστικού.

```
conflict_latent_scores<-factor.scores.ltm(conflict_2PL, method="EAP", resp.patterns
= conflict)
summary(conflict_latent_scores$score.dat$z1)
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -1.3918699 -0.6610675 -0.0225900  0.0001959  0.5882801  2.0068550
sd(conflict_latent_scores$score.dat$z1)
## [1] 0.8461795
```

Από τα αποτελέσματα φαίνεται η συμφωνία των δύο προγραμμάτων IRTPRO και ltm σε ακρίβεια τετάρτου δεκαδικού ψηφίου.

**EXAMINING INSTRUMENTS' PSYCHOMETRIC  
PROPERTIES WITHIN THE ITEM RESPONSE  
THEORY FRAMEWORK:  
FROM THEORY TO PRACTICE**

*Nikolaos Tsigilis*

*Aristotle University of Thessaloniki, Greece*

**Abstract:** Item Response Theory (IRT) represents an alternative to Classical Test Theory (CTT) framework for developing new instruments and examining the psychometric properties of existing ones. IRT models the relationship between a latent trait and the response of each item of a measurement tool. Thus, contrary to CTT, the focus of IRT is on the item level. In the present study the basic concepts of IRT are presented along with its advantages in relation to CTT, the assumptions underlying its implementation and how the results can be interpreted and utilized in an attempt to sensitize researchers for its applicability and usefulness. The study focuses on the simplest models where dichotomous data are used as their in-depth understanding facilitates a smooth transition to the application of more complex IRT techniques. Finally, an example of IRT analysis is presented in detail using two statistical programs, a commercial one with graphical environment (IRTPRO) and an open source (ltm) in the R environment.

**Key words:** Item Response Theory, Latent trait models, Test psychometric properties

**Address:** Dr Nikolaos Tsigilis, Department of Journalism & Mass Media Communication, Aristotle University of Thessaloniki, 46 Egnatia Ave, 54625 Thessaloniki, Greece. Tel: +30-2310-991997. E-mail: ntsigilis@jour.auth.gr