

Θεωρητικές έννοιες
Μετρικής και Ψυχομετρίας

Σημειώσεις του μαθήματος

Ψυχομετρία I

K. A. Μυλωνάς

Τρίτη έκδοση, 5^η εκτύπωση
Αθήνα, 2007

Πρόλογος

Οι σημειώσεις αυτές αποτελούν, κατά το εφικτό, μια πρώτη γνωριμία με τη θεωρητική ψυχομετρία, δηλαδή με τις μεθόδους και τις έννοιες εκείνες που συνθέτουν την προσπάθεια για την καλύτερη, πληρέστερη και, κυρίως, ασφαλέστερη μέτρηση των διαφόρων χαρακτηριστικών του ανθρώπου. Με την συνδρομή της Πειραματικής Ψυχολογίας και της Ψυχολογίας Ατομικών Διαφορών και στηριζόμενοι στις δυνατότητες της Ερευνητικής Μεθοδολογίας και της Εφαρμοσμένης στην Ψυχολογία Στατιστικής, οι ψυχομέτρες έχουν επιχειρήσει, τις τελευταίες δεκαετίες, να δημιουργήσουν αξιόπιστα και έγκυρα ψυχοτεχνικά μέσα για τη μέτρηση των ψυχολογικών ιδιοτήτων και των μη-έκδηλων στη συμπεριφορά χαρακτηριστικών της προσωπικότητας, της νόησης, των ικανοτήτων, των προτιμήσεων, στάσεων, απόψεων, κ.ο.κ. Οι προσπάθειες αυτές δεν είναι κάτι νέο και στην ελληνική πραγματικότητα, είναι όμως μεγάλη, επίσης, η ανάγκη για περαιτέρω και περισσότερο συστηματική αποδοτικότητα στη δημιουργία και στον έλεγχο τέτοιων ψυχομετρικών εργαλείων και στη χώρα μας. Η ικανότητα εκτίμησης των δυνατοτήτων που κάθε νέο (ή και παλαιότερο) ψυχομετρικό τεστ προσφέρει θα πρέπει, επίσης, να αποτελεί κεκτημένο πεδίο γνώσεων του νέου κοινωνικού επιστήμονα, καθώς καθημερινή πλέον είναι η ανάγκη χρήσης των συμπερασμάτων των ερευνών που στηρίζονται σε τέτοια ψυχομετρικά εργαλεία.

Οι παρούσες σημειώσεις –στην τρίτη έκδοσή τους– φιλοδοξούν, με τη συνδρομή και άλλων πηγών, να εντοπίσουν και να πραγματευθούν, σε εισαγωγικό επίπεδο, μερικές από τις βασικές έννοιες και τεχνικές που χρησιμοποιεί η σύγχρονη Ψυχομετρία. Είναι προφανές ότι οι σημειώσεις αυτές αποτελούν άρρηκτη συνέχεια και συμπλήρωμα των αντίστοιχων Πανεπιστημιακών διαλέξεων.

Σεπτέμβριος 2007,

Κ. Α. Μυλωνάς

ΠΕΡΙΕΧΟΜΕΝΑ

Μέρος 1. Εισαγωγή – Βασικές έννοιες στην Ψυχομετρία

Ατομικές διαφορές: επίκεντρο και έναυσμα

Ψυχομετρία: Από τη θεωρία στην επιστήμη

Μέρος 2. Κλίμακες και σφάλματα μέτρησης

Η σχολή της λειτουργικότητας και η σχολή των ιδιοτήτων

Κατηγορικές και Τακτικές Κλίμακες Μέτρησης

Κλίμακες Ίσων Διαστημάτων

Κλίμακες Ίσων Λόγων

Η θεωρία των αληθών τιμών

Μέρος 3. Αξιοπιστία μετρήσεων και αξιοπιστία των τεστ

Γενικές έννοιες

Αξιοπιστία και σφάλματα μέτρησης

Δείκτης αξιοπιστίας επαναληπτικών μετρήσεων

Αξιοπιστία εναλλακτικών τύπων ή ισοδύναμων τύπων

Αξιοπιστία των ημικλάστων

Αξιοπιστία εσωτερικής συνέπειας

Μεταξύ βαθμολογητών αξιοπιστία

Μέρος 4. Εγκυρότητα μετρήσεων και εγκυρότητα των τεστ

Γενικές έννοιες

Κατά τεκμήριο εγκυρότητα

Εγκυρότητα αντιπροσωπευτικού περιεχομένου

Εγκυρότητα προβλεπτικής συνάφειας ή προβλεπτική εγκυρότητα

Εγκυρότητα συγχρονικής συνάφειας

Εγκυρότητα εννοιολογικής δομής

Μέρος 5. Πολώσεις

Είδη πολώσεων

Ανίχνευση των πολώσεων ερωτημάτων

Μέρος 6. Έλεγχος ερωτημάτων

Ανάλυση Ερωτημάτων

Προϋποθέσεις

Στόχοι της ανάλυσης ερωτημάτων

Διαδικασία ανάλυσης ερωτημάτων

Αποτελέσματα και περαιτέρω διαδικασία

Κριτήρια και λάθη στην τελική επιλογή ερωτημάτων

Ανάλυση Παραγόντων

Προϋποθέσεις

Μαθηματική βάση της ανάλυσης παραγόντων

Αποτελέσματα ανάλυσης κυρίων συνιστωσών και αποτελέσματα ανάλυσης παραγόντων

Μέρος 7. Λανθάνοντα χαρακτηριστικά (*latent traits* - 'θ')

Θεωρία των Λανθανόντων Χαρακτηριστικών (Item Response Theory)

Το μοντέλο Rasch

Μέρος 8. Μετατροπή των αρχικών τιμών σε δευτερογενείς κλίμακες

Μέρος 1. Εισαγωγή – Βασικές έννοιες στην Ψυχομετρία

Ατομικές διαφορές: επίκεντρο και έναυσμα

Η γενική ψυχολογική επιστημονική διαδικασία, η μεθοδολογική προσέγγιση και η ερευνητική δραστηριότητα, καθώς και οι ανάγκες για όλο και μεγαλύτερη και πληρέστερη γνώση των ανθρώπινων χαρακτηριστικών με σκοπό τη διασάφηση των παραμέτρων που διέπουν και καθορίζουν την ανθρώπινη συμπεριφορά, είναι μερικοί μόνον από τους λόγους που οδήγησαν στην ανάπτυξη του κλάδου της Ψυχομετρίας στα πλαίσια της ψυχολογικής ερμηνείας των φαινομένων, της μέτρησης των ανθρώπινων χαρακτηριστικών και της παρέμβασης των ειδικών, όπου αυτή θεωρείται αναγκαία.

Η μέτρηση των ψυχολογικών χαρακτηριστικών και παραμέτρων που αφορούν στον άνθρωπο απετέλεσε, πρώτιστα, αντικείμενο μελέτης της *Ψυχολογίας των Ατομικών Διαφορών* και εδώ θα πρέπει να αναζητηθεί η επιστημονική βάση της Ψυχομετρίας. Στην ουσία, οι ατομικές διαφορές (οι οποίες μπορεί να είναι διατομικές, διομαδικές ή και ενδοατομικές) αποτελούν, ως γνωστόν, τη βάση για το σύνολο της επιστήμης της ψυχολογίας. Οι διαφορές αυτές αποτελούν φυσικό νόμο, καθώς κανείς μας δεν είναι απόλυτα όμοιος με κάποιον άλλον και θα ήταν τουλάχιστον παράδοξο να υπάρξουν έστω και δύο πανομοιότυπες “προσωπικότητες”. Ο σκοπός της παρουσίασης βέβαια, δεν είναι τα ηθικά ζητήματα που προκύπτουν αλλά η μελέτη των υπαρχόντων ατομικών διαφορών με συστηματικό και συμβατό τρόπο. Όσον αφορά στα ανθρώπινα χαρακτηριστικά, τίποτε δεν είναι απόλυτο ή σταθερό όπως συμβαίνει, για παράδειγμα, στην επιστήμη της Φυσικής. Στην επιστήμη της Ψυχολογίας δεν υπάρχουν απόλυτες συναρτήσεις, διότι σε τέτοιου είδους μαθηματικές απεικονίσεις δεν υπάρχουν περιθώρια σφάλματος. Στην Ψυχολογία, οι συναρτήσεις πρόβλεψης ή επεξήγησης της ανθρώπινης συμπεριφοράς περιλαμβάνουν όλες εκείνες τις παραμέτρους που μπορεί να καθορίζουν τη συμπεριφορά (και οι οποίες μπορεί να είναι λίγες ή και πολλές σε αριθμό) αλλά πάντα περιλαμβάνουν και την παράμετρο e (error) η οποία και αντιπροσωπεύει το σφάλμα που οφείλεται στις ατομικές διαφορές και άλλα είδη σφαλμάτων.

Η παράμετρος e δεν είναι σταθερή και, στις περισσότερες περιπτώσεις, δεν μπορεί να ερμηνευθεί ως το αποτέλεσμα κάποιων συγκεκριμένων αιτίων. Συνήθως, η παράμετρος αυτή οφείλεται στην τυχαιότητα, δηλαδή οφείλεται στο γεγονός ότι κανείς μας δεν ταυτίζεται ως προς τα χαρακτηριστικά του με κάποιον άλλον. Στη *Στατιστική*, η τυχαιότητα αυτή είναι στην πραγματικότητα η διακύμανση (αλλιώς, διασπορά) που δεν οφείλεται μόνο σε συγκεκριμένες συστηματικές αιτίες ή παραμέτρους, αλλά οφείλεται, τουλάχιστον εν μέρει, στην τυχαία δειγματοληψία κάποιας ερευνητικής διαδικασίας¹. Θα πρέπει να διευκρινισθεί ότι η διακύμανση αυτή, επειδή ακριβώς οφείλεται στις ατομικές διαφορές που παρατηρούνται μέσα σε ένα σύνολο ατόμων, δεν είναι «άχρηστη», ούτε και θα πρέπει να αγνοηθεί. Πολλές φορές, δεν μπορούμε να την ελέγξουμε ή να την περιορίσουμε, αλλά είναι επιτακτική η ανάγκη της επεξήγησης και ερμηνείας όσο το δυνατόν μεγαλύτερου τμήματός της.

Με βάση τις εκτεταμένες και πολυδιάστατες έρευνες που έχουν διεξαχθεί τον 20^ο αιώνα και μέχρι σήμερα, έχουμε καταλήξει, χωρίς ποτέ να φτάσουμε στην απόλυτη πραγματικότητα, σε θεωρητικές προσεγγίσεις και πρότυπα επεξήγησης της ανθρώπινης συμπεριφοράς μέσω διαφόρων παραμέτρων. Γνωρίζοντας, επομένως τις παραμέτρους αυτές, μπορούμε να αποφανθούμε για τα χαρακτηριστικά εκείνα που διαφοροποιούν τον κάθε άνθρωπο από τους άλλους, να περιγράψουμε δηλαδή **διαφορικά** τον κάθε άνθρωπο. Η διαφορική αυτή προσέγγιση είναι βέβαια πολύπλευρη και μπορεί να αναφέρεται στην περιγραφή των χαρακτηριστικών αυτών σε σχέση με τον υπόλοιπο πληθυσμό, κάτι που θα μας δώσει πληροφορίες για πιθανές αποκλίσεις του συγκεκριμένου ανθρώπου ως προς τα χαρακτηριστικά αυτά· μπορεί, επίσης, να αναφέρεται, μεταξύ άλλων, στη σχέση των χαρακτηριστικών αυτών με άλλα χαρακτηριστικά του ίδιου ανθρώπου, με σκοπό την πολυδιάστατη ενδοατομική επεξήγηση της ανθρώπινης συμπεριφοράς.

Σε κάθε περίπτωση, με όποιον τρόπο και αν προσεγγίζουμε την ανθρώπινη οντότητα, είναι βέβαιο ότι παρεμβάλλονται ανεξέλεγκτοι παράγοντες οι οποίοι και αποκρύπτουν μέρος της «αλήθειας», ενώ ταυτόχρονα είναι ήδη γνωστό πως τα θεωρητικά πρότυπα που στηρίζουν την προσέγγιση αυτή περιλαμβάνουν την

¹ Περισσότερες πληροφορίες επί του θέματος μπορούν να αναζητηθούν στα συγγράμματα:

I.N. Παρασκευόπουλος (1990), *Στατιστική εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμος Β' Επαγωγική Στατιστική, Κεφ. 8.1. και 8.2. και,

παράμετρο της τυχαιότητας. Επομένως τα σφάλματα υφίστανται και λόγω της τυχαιάς δειγματοληψίας αλλά και λόγω των συστηματικών επιδράσεων διαφόρων εξωγενών παραμέτρων.

Συνοψίζοντας, δύο είναι τα κύρια σημεία ενδιαφέροντος: **α)** το γεγονός ότι θέλουμε να είμαστε σε θέση να μετρήσουμε τις ανθρώπινες ιδιότητες και να τις εντάξουμε στο γενικότερο πλαίσιο των ατομικών διαφορών ώστε να είναι ερμηνεύσιμες και, **β)** οι μετρήσεις αυτές να είναι στο μέγιστο δυνατό βαθμό **ακριβείς**, δηλαδή να ερμηνεύσουμε, όπως έχει ήδη συζητηθεί, όσο περισσότερο γίνεται την παράμετρο *e*. Η αναγκαιότητα ακρίβειας στις μετρήσεις οδήγησε στην προσπάθεια για δημιουργία και ανάπτυξη συστηματικών διαδικασιών που θα εξασφάλιζαν, όσο το δυνατόν, την ακρίβεια αυτή.

Ψυχομετρία: Από τη θεωρία στην επιστήμη

Στο *Chambers Twentieth-Century Dictionary*, η Ψυχομετρία ορίζεται με δύο τρόπους: **α)** ο κλάδος της ψυχολογίας που ασχολείται με μετρήσιμους παράγοντες και, **β)** η μεταφυσική ικανότητα του να καθορίζει κανείς τα χαρακτηριστικά των πραγμάτων μέσω της απλής επαφής μαζί τους. Ασφαλώς δεν μας ενδιαφέρει η μεταφυσική πλευρά της ψυχομετρίας, πλευρά που δεν είναι βέβαια ανάξια λόγου και διαδραματίζει ιδιαίτερα σημαντικό ρόλο στην ψυχολογία των καθημερινών καταστάσεων (βλέπε και ‘Θεωρία της Ψυχολογίας του αφελούς’ κατά *Heider*, 1958), μας ενδιαφέρουν, όμως, κυρίως οι μετρήσιμοι αυτοί παράγοντες οι οποίοι αντιπροσωπεύουν όλους τους κλάδους της ψυχολογικής έρευνας για την ανθρώπινη συμπεριφορά. Για κάθε τέτοια συμπεριφορά, είναι δυνατό να δημιουργηθεί και να χορηγηθεί ένα ψυχομετρικό εργαλείο που ως κύριο στόχο του θα έχει τη διερεύνηση του αντίστοιχα μετρήσιμου παράγοντα σε ατομική ή διατομική βάση.

Θα πρέπει να τονίσουμε ότι η ανάπτυξη της ψυχομετρίας οφείλεται, ως προς τους εν λόγω μετρήσιμους παράγοντες, στην προσπάθεια μέτρησης της νοημοσύνης (*Galton, Spearman, Guilford, Pearson, Binet, Eysenck, Wechsler, Sternberg, κ.ά.*), που αποτελούσε από την απαρχή της επιστήμης της ψυχολογίας ιδιαίτερα σημαντικό ερώτημα. Αρκεί να θυμηθούμε τις ερευνητικές προσεγγίσεις που αφορούν σε

διδύμους, που ακόμη και σήμερα αποτελούν αντικείμενο συζήτησης και διαφωνίας. Θα πρέπει όμως επίσης να τονίσουμε ότι ο στόχος της ψυχομετρίας δεν είναι να μετρήσει **μόνο** τη νοημοσύνη, αλλά ολόκληρο το φάσμα της ανθρώπινης οντότητας. Έτσι, ένας άλλος μεγάλος τομέας στον οποίο έχει αφιερώσει η ψυχομετρία ιδιαίτερες προσπάθειες είναι η προσωπικότητα, και ακολουθεί ο τομέας των προσωπικών ικανοτήτων και δεξιοτήτων, ο τομέας των ενδιαφερόντων, στάσεων και αξιών, ο τομέας επιτευγμάτων κατά την εκπαιδευτική διαδικασία, ο τομέας της κλινικής διάγνωσης και συμβουλευτικής, καθώς και ο τομέας της επαγγελματικής συμβουλευτικής, κ.ά. Είναι προφανές ότι η δημιουργία ψυχομετρικών εργαλείων και η προσέγγιση των μετρήσιμων παραγόντων στους οποίους τα τεστ αυτά αναφέρονται, καλύπτει κάθε θεωρητική πτυχή και επηρεάζει κάθε ψυχολογική πρακτική και εφαρμογή.

Κατά πόσο όμως μπορούμε να υποστηρίξουμε ότι αυτή η προσέγγιση των «μετρήσιμων παραγόντων» είναι αποτέλεσμα επιστημονικής προσπάθειας; Κατά πόσο δηλαδή διαφοροποιείται η λογική της ψυχομετρίας από αυτήν της μεταφυσικής της ιδιότητας; Ο *Wittgenstein* το 1972 μίλησε για «εννοιολογική σύγχυση» (*conceptual confusion*) στους κόλπους της ψυχολογίας και υποστήριξε ότι εκεί οφείλονται οι περιορισμοί της επιστημονικής ταυτότητας της ψυχολογίας, εν γένει. Από την άλλη, οι λεγόμενες «σκληρές» επιστήμες (Φυσική, Χημεία, Μαθηματικά, κ.λπ.), από την αρχή μάλιστα του αιώνα, διαφοροποιούνταν από την ψυχολογία και υπερτόνιζαν την επιστημονική τους βαρύτητα σε σχέση με την συγχυτική δραστηριότητα της ψυχολογίας. Με τη λογική αυτή, η ψυχολογία δεν φαινόταν να είναι σε θέση να αναπτύξει την επιστήμη της μέτρησης των ανθρώπινων ιδιοτήτων, πράγμα που και οι ίδιοι οι ψυχολόγοι διαπίστωναν και αναγνώριζαν. Το «σύμπλεγμα» αυτό της ψυχολογίας έθετε φραγμούς στην αποδεικτική ικανότητα και έπρεπε να λυθεί. Μία λύση έδωσαν κάποιοι πρωτοπόροι ερευνητές και θεωρητικοί στην δεκαετία του 1970 άρχισαν να σκέπτονται ως προς τη γένεση μιας καινούριας επιστήμης, της «Γνωστικής Επιστήμης». Η γένεση αυτή είχε σκοπό τη διεπιστημονική συνένωση δυνάμεων ώστε να «χαρτογραφηθεί» ο τρόπος της ανθρώπινης σκέψης. Για τη χαρτογράφηση αυτή, χρησιμοποιήθηκαν ευρύτατα οι Ηλεκτρονικοί Υπολογιστές της εποχής, γεγονός που οδήγησε στην ανάπτυξη νέων επιστημών, όπως αυτή της Τεχνητής Νοημοσύνης. Με τη χρήση όμως των

υπολογιστών, ενέσχυσε και η ανάγκη της όλο και ισχυρότερης και πληρέστερης επεξεργασίας των πληροφοριών, γεγονός που οδήγησε στην εξέλιξη των υπολογιστικών διαδικασιών και στην ανάπτυξη εξειδικευμένου και ισχυρότατου λογισμικού. Η νέα αυτή δυνατότητα έγινε ευρύτατα αποδεκτή και ταυτόχρονα επέτρεψε στην Μαθηματική Ψυχολογία να ξεπεράσει τα πρακτικά εμπόδια και να εξελίξει τα μαθηματικά υποδείγματα επεξεργασίας πληροφοριών. Η παράλληλη ανάπτυξη της Μεθοδολογίας Επιστημονικής Έρευνας στην ψυχολογία και οι νέες στατιστικές μέθοδοι επέτρεψαν στους ψυχολόγους να ισχυριστούν ότι διαθέτουν πλέον αποδεικτικά εργαλεία για τις θεωρίες και τις παρατηρήσεις τους. Όλο το πλέγμα των διαδικασιών αυτών επέτρεψε και τη σαφέστερη και ειδική μελέτη των μετρήσιμων ανθρώπινων παραγόντων. Αξίζει να σημειωθεί ότι εάν σήμερα διαθέτουμε το MMPI (*Minnesota Multiphasic Personality Inventory*), αυτό οφείλεται στη δυνατότητα επεξεργασίας πολυμεταβλητών χαρακτηριστικών της προσωπικότητας, δηλαδή στη δυνατότητα στατιστικής συνεξέτασης περισσότερων από μιας μετρήσιμης παραμέτρου, κάτι που δεν θα ήταν εφικτό χωρίς μεθοδολογική και στατιστική ισχύ. Το ίδιο συμβαίνει και για το ITPA (*Illinois Test of Psycholinguistic Abilities*) καθώς και για πολλά άλλα πολυδιάστατα τεστ. Το παραπάνω σκεπτικό δεν σημαίνει ότι πριν την ανάπτυξη της γνωστικής επιστήμης και των όσων επακολούθησαν δεν είχαμε εξέλιξη της στατιστικής και της μεθοδολογίας, ή ότι οι ερευνητές και οι θεωρητικοί δεν διέθεταν επιστημονικές ικανότητες, αλλά απλά τονίζει τα άλματα τα οποία παρατηρήθηκαν στα τελευταία 30 περίπου χρόνια.

Το σημαντικότερο ίσως κέρδος από τα άλματα αυτά ήταν ότι η ψυχολογία ξεπέρασε το σύμπλεγμα της «θεωρητικής επιστήμης» που της είχε επισυναφθεί –είτε δίκαια είτε άδικα– και απέκτησε εμπιστοσύνη στις αποδεικτικές δυνατότητές της. Άσχετα με το αν τελικά έχουμε σήμερα τέτοιες δυνατότητες ως ψυχολόγοι, αυτό που τουλάχιστον μπορούμε να ισχυριστούμε είναι ότι στις περισσότερες περιπτώσεις, όποιες παραμέτρους της ανθρώπινης οντότητας προσεγγίζουμε, τουλάχιστον τις μετρούμε με αρκετά ικανοποιητική ακρίβεια και ότι τα συμπεράσματά μας δεν είναι παρακινδυνευμένα. Έχουμε λοιπόν κατορθώσει να ξεπεράσουμε το εμπόδιο της συγχυτικής εννοιολογικής πληθώρας και προσπαθούμε να επισυνάψουμε σε κάθε μία μετρήσιμη παράμετρο διευκρινίσεις ως προς τη λειτουργία και τις ιδιότητές της. Όμως, αν παραμείνουμε στο στάδιο αυτό, χωρίς να υπολογίσουμε την πρακτική

χρησιμότητα των ανακαλύψεών μας αυτών και δεν χρησιμοποιήσουμε την αποδεικτική ισχύ των νέων μεθόδων που διαθέτουμε για να δώσουμε λύσεις σε καίρια θέματα που απασχολούν την ψυχολογική πρακτική, τότε έχουμε αποτύχει. Η «παγίδα» αυτή της επιμονής στη μετρική μέθοδο και της διεξοδικής περιγραφής των μετρικών χαρακτηριστικών των τεστ, χωρίς εστίαση στο στόχο που οι μετρήσεις αυτές θα εξυπηρετούσαν, ήταν πολύ ισχυρή για αρκετά μεγάλο χρονικό διάστημα. Οι Cronbach, Gleser, Nanda, και Rajaratnan², με τη θεωρία της γενικευσιμότητας (*generalizability theory*) άσκησαν δριμεία κριτική ως προς το θέμα αυτό (βλ. και Πρόλογο του A. Kaufman στο Georgas, Weiss, van de Vijver, & Saklofske (Eds.) *Culture and Children's Intelligence: Cross-cultural analysis of the WISC-III*, υπό έκδοση, Elsevier Science). Ας μη λησμονούμε ότι ο κύριος στόχος δεν είναι η μέτρηση και μόνο των χαρακτηριστικών και ιδιοτήτων του ανθρώπου, αλλά η ακριβής γνώση των παραμέτρων αυτών ώστε να είμαστε σε θέση να προχωρήσουμε σε περαιτέρω ψυχολογική πράξη.

Μέρος 2. Κλίμακες και σφάλματα μέτρησης

Η σχολή της λειτουργικότητας και η σχολή των ιδιοτήτων

Υπάρχουν δύο ευρεία θεωρητικά πλαίσια μέσα στα οποία μπορεί κανείς να δημιουργήσει ένα ψυχομετρικό εργαλείο. Για το πρώτο πλαίσιο, *τη σχολή της λειτουργικότητας*, ο,τιδήποτε και αν μετρήσουμε έχει νόημα μόνο και μόνο ως προς την εφαρμογή που θα ακολουθήσει. Η θεωρητική αυτή ψυχομετρική προσέγγιση απετέλεσε έντονο έναυσμα, τα τελευταία ειδικά χρόνια, για την δημιουργία πληθώρας εργαλείων μέτρησης καθώς οι πρακτικές ανάγκες ολοένα και μεγαλώνουν. Από την άλλη, η θεωρητική προσέγγιση που αφορά στις ατομικές ιδιότητες και χαρακτηριστικά, έχει τη βάση της κυρίως στην αναζήτηση της προσωπικότητας του ατόμου. Στην αναζήτηση αυτή δεν υπάρχουν απόλυτες τιμές, αλλά ένα συνεχές στο οποίο εμπίπτουν όλες οι διαφοροποιήσεις μεταξύ των ατόμων. Με άλλα λόγια, δεν υπάρχουν μόνο τα δύο άκρα του συνεχούς, αλλά ολόκληρο το συνεχές. Όπως είναι

² Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnan, N. (1972). *The Dependability of Behavioral Measurements:*

προφανές, η θεωρητική προσέγγιση της σχολής της λειτουργικότητας διαφοροποιεί περισσότερο ποιοτικά, ενώ η προσέγγιση των χαρακτηριστικών διαφοροποιεί περισσότερο ποσοτικά. Οι διαφοροποιήσεις αυτές είναι θεωρητικού τύπου και δεν αναφέρονται στην τελική αριθμητική έκφρασή τους, παρά μόνο στον τρόπο προσέγγισης των ίδιων παραμέτρων με διαφορετικό τρόπο. Τι είναι όμως ποιοτική και τι είναι ποσοτική διαφοροποίηση; Σε τι διαφέρουν από την «αριθμοποίηση»; Ποιους τρόπους διαθέτουμε στην ψυχομετρία για να διαφοροποιούμε αρχικά μεταξύ των τιμών και στη συνέχεια να τις «αριθμοποιούμε»;

Κατηγορικές και Τακτικές Κλίμακες Μέτρησης

Κάθε έννοια που μεταβάλλεται από άτομο σε άτομο ή μεταξύ καταστάσεων και συνθηκών ή ακόμη και χωρο-χρονικά, χρειάζεται, για να μετρηθεί και να εκφραστεί αριθμητικά, ένα κοινό μέτρο σύγκρισης, ένα σύστημα ταξινόμησης και οριοθέτησης των τιμών που είναι δυνατόν να προκύψουν. Ανάλογα με το είδος των μεταβλητών στις οποίες αναφερόμαστε, υπάρχουν και διαφορετικά τέτοια συστήματα ταξινόμησης, ή κλίμακες. Τέτοιες κλίμακες μέτρησης είναι οι κατηγορικές ή ονοματικές (*nominal scale*) οι οποίες και αναφέρονται σε μεταβλητές με δύο ή περισσότερες σαφώς διαχωρισμένες και μη-διατάξιμες πιθανές τιμές, δηλαδή τιμές που δεν μπορούν να ιεραρχηθούν με βάση μια λογική σειρά. Το κλασικό παράδειγμα είναι αυτό της μεταβλητής του «φύλου» όπου οι δύο πιθανές, σαφώς διαχωρισμένες τιμές είναι «άνδρας» και «γυναίκα». Οι δύο αυτές τιμές δεν είναι δυνατόν να ιεραρχηθούν σε αύξουσα ή φθίνουσα σειρά, καθώς, αν αντιστρέψουμε τη σειρά αναφοράς στις τιμές αυτές, δεν αλλάζει τίποτε. Οι ονοματικές λοιπόν κλίμακες απλά επισυνάπτουν «τίτλους» στις ενδεχόμενες διαφοροποιήσεις της μεταβλητής. Ένα ακόμη παράδειγμα είναι αυτό της εθνικότητας: «Έλληνας», «Γερμανός», «Κινέζος», κ.ο.κ.

Όταν οι τιμές στις οποίες θα καταλήξουμε χρησιμοποιώντας την κλίμακα μέτρησης της μεταβλητής είναι διατάξιμες, δηλαδή, όταν μπορούμε να τις ιεραρχήσουμε σε αύξουσα ή φθίνουσα σειρά, τότε οι κλίμακες που χρησιμοποιούμε είναι οι τακτικές κλίμακες μέτρησης (*ordinal scale*). Για παράδειγμα, ως προς τη

μεταβλητή «εμβαδόν αιθουσών της Φιλοσοφικής Σχολής», θα μπορούσαμε, μετρώντας με μια τακτική κλίμακα μέτρησης, να καταλήξουμε σε τρεις τιμές: μεγάλες, μεσαίες και μικρές ως προς το εμβαδόν αίθουσες. Θα μπορούσαμε να τις διατάξουμε και αντίστροφα: μικρές, μεσαίες, μεγάλες ως προς το εμβαδόν αίθουσες. Σε καμία όμως περίπτωση δεν θα μπορούσαμε να τις διατάξουμε ως μικρές, μεγάλες και μεσαίες αίθουσες, διότι τότε καταστρατηγούμε τη φυσική διάταξη των πραγμάτων, στερούμεθα πληροφοριών και καθιστούμε τη μέτρησή μας στην ουσία προβληματική ως άχρηστη. Ένα άλλο, πολύ συχνό, παράδειγμα, είναι αυτό του κοινωνικο-οικονομικού επιπέδου και ένα τρίτο, αυτό των τάξεων του δημοτικού σχολείου (Α΄ ως Στ΄).

Για πρακτικούς και στατιστικούς κυρίως λόγους, για τις τακτικές κλίμακες μέτρησης και κυρίως για εκείνες που περιλαμβάνουν αρκετές, έως πολλές, διαβαθμίσεις μπορούμε να επισυνάψουμε έναν αριθμό σε αύξουσα σειρά στην κάθε διαβάθμιση. Τέτοιες κλίμακες μέτρησης είναι οι γνωστές *Likert* κλίμακες 4, 5, 6 ή 7 βαθμίδων, οι κλίμακες *Thurstone*, κ.ά. Οι βαθμίδες των κλιμάκων αυτών είναι, στην ουσία, ποιοτικές διαβαθμίσεις ενός συνεχούς. Οι αποστάσεις μεταξύ ζευγών των βαθμίδων αυτών δεν είναι ποτέ ίσες: η βαθμίδα «πολύ», ως βαθμίδα, διαφέρει έστω X ποσότητα από τη βαθμίδα «μέτρια» ενώ η βαθμίδα «λίγο» διαφέρει Ψ ποσότητα από τη βαθμίδα «μέτρια». Επίσης, οι αποστάσεις αυτές δεν είναι σταθερές, αλλά εξαρτώνται από την αντίληψη που κάθε άτομο έχει για τις διαβαθμίσεις αυτές. Όμως, καθώς αυτές οι μετρήσεις μπορούν να αριθμοποιηθούν («Πάρα πολύ» = 5 έως «Καθόλου» = 1), χρησιμοποιούνται –κάτω από συνθήκες και προϋποθέσεις– στατιστικώς ως ποσοτικές *εκτιμήσεις* της οποιασδήποτε έννοιας μετρά η συγκεκριμένη κλίμακα.

Κλίμακες Ίσων Διαστημάτων

Οι κλίμακες αυτές αναφέρονται σε μεταβλητές που οι βαθμίδες μέτρησής τους ισαπέχουν αριθμητικά. Η θερμοκρασία μπορεί να μετρηθεί με κλίμακα ίσων διαστημάτων καθώς για κάθε βαθμό Κελσίου (ή Φαρενάιτ) η αριθμητική διαφοροποίηση μεταξύ των βαθμίδων είναι ακριβώς όμοια. Όμως, πρέπει να γίνει η ακόλουθη επισήμανση: τα ίσα πηλικά βαθμίδων (π.χ. $20^\circ \div 10^\circ = 2$ και $40^\circ \div 20^\circ = 2$) δεν είναι εννοιολογικά ίσα, δηλαδή δεν είναι **διπλάσια** η θερμοκρασία μεταξύ των

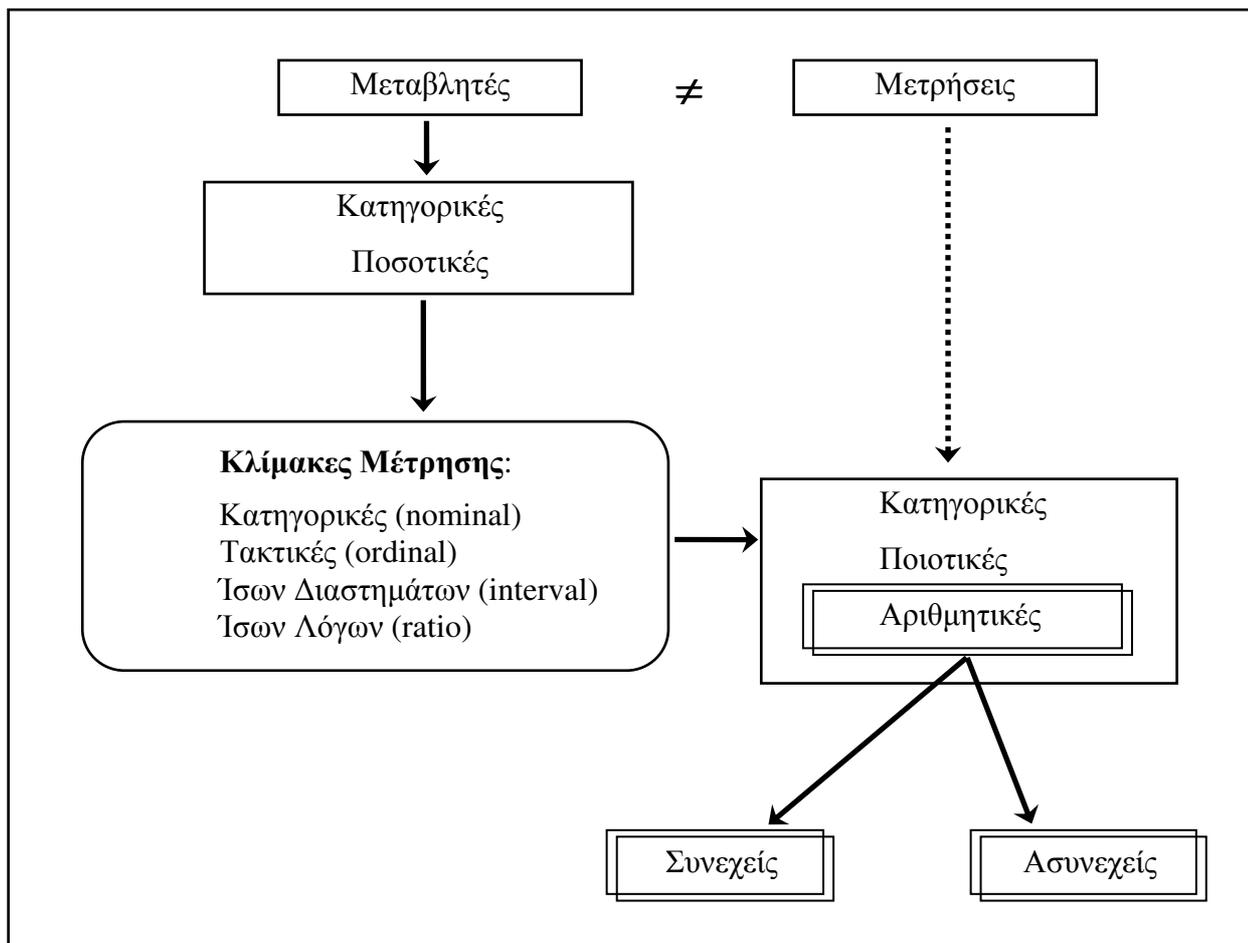
δύο βαθμίδων και στις δύο περιπτώσεις. Βέβαια, οι κλίμακες ίσων διαστημάτων καταλήγουν σε ίσες αριθμητικά διαφοροποιήσεις, αλλά οι αριθμητικά ίσοι λόγοι μεταξύ των βαθμίδων δεν είναι όμοιοι ως προς την ερμηνεία τους. Ένα άλλο παράδειγμα είναι η μέτρηση ψυχολογικών παραμέτρων όπως το άγχος. Με τη μέθοδο αξιολόγησης Taylor κάποιος θα μπορούσε να συγκεντρώσει τελικό δείκτη 34. Αυτό δεν σημαίνει ότι το άγχος του είναι διπλάσιο από κάποιον άλλον που συγκεντρώνει τελικό δείκτη άγχους 17, ενώ το ίδιο συμβαίνει και για δύο άλλα πρόσωπα που έχουν βαθμούς άγχους 20 και 10 αντίστοιχα. Η αριθμητική διαφορά μεταξύ των βαθμίδων άγχους είναι πάντα μία μονάδα, όμως οι διαφοροποιήσεις στα διάφορα σημεία του συνεχούς *δεν είναι ανάλογες*.

Κλίμακες Ίσων Λόγων

Οι κλίμακες αυτές μέτρησης (*ratio scale*) αναφέρονται σε καθαρά ποσοτικές διαφοροποιήσεις, οι οποίες και εκφράζονται με αριθμητικές τιμές που απέχουν ισόποσα μεταξύ τους. Αν ένα άτομο ζυγίζει 54 κιλά και ένα άλλο άτομο 57, τότε η διαφορά τους είναι 3 κιλά βάρους. Η ίδια διαφορά ισχύει και για δύο άλλα άτομα τα οποία ζυγίζουν 78 και 81 κιλά αντίστοιχα. Από την κλίμακα αυτή μέτρησης προκύπτουν ασυνεχείς ή συνεχείς τιμές (ακέραιοι ή πραγματικοί $\{\mathfrak{R}\}$ αριθμοί, συνήθως θετικού προσήμου), οι οποίοι μπορεί να έχουν πεπερασμένο εύρος ή όχι. Η ουσιαστική διαφορά με τις κλίμακες ίσων διαστημάτων είναι ότι κάποιος που ζυγίζει 50 κιλά είναι δύο φορές πιο βαρύν από κάποιον που ζυγίζει 25 κιλά και το ίδιο συμβαίνει για κάποιον που ζυγίζει 100 κιλά σε σχέση με κάποιον που ζυγίζει 50 κιλά.³

Το ακόλουθο σχήμα παρουσιάζει συνοπτικά τη σύνδεση μεταβλητών και μετρήσεων μέσω της χρήσης των κλιμάκων μέτρησης:

³ Για τις έννοιες των μεταβλητών και των μετρήσεων, βλέπε Κεφάλαια 1.2. και 1.3., στο σύγγραμμα **I.N. Παρασκευόπουλος** (1990), *Στατιστική εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμος Α΄ Περιγραφική Στατιστική.



Ανακεφαλαιώνοντας, οι κλίμακες μέτρησης αποτελούν περισσότερο ερευνητική διευκόλυνση και αφορούν στην συλλογή δεδομένων κατά την ερευνητική διαδικασία. Η λογική τους, όμως, είναι και η βάση της ψυχομετρικής πρακτικής. Οι κλίμακες μέτρησης αντιπροσωπεύουν το θεωρητικό υπόβαθρο των αρχικών τουλάχιστον μετρήσεων του οποιουδήποτε ψυχομετρικού εργαλείου. Στο σημείο αυτό, θα πρέπει να διαφοροποιήσουμε την έννοια της ψυχομετρικής κλίμακας αξιολόγησης από τις κλίμακες μέτρησης. Η ψυχομετρική κλίμακα αξιολόγησης είναι το ίδιο το ψυχομετρικό εργαλείο, ή τμήματά του. Η κλίμακα αυτή έχει δημιουργηθεί ως νέο είδος γνώμονα, που έχει υποκειμενικό χαρακτήρα και ενώ μπορεί να αριθμοποιηθεί, στην ουσία εκφράζει ποιότητες.

Αν, για παράδειγμα, θέλουμε να μετρήσουμε τη μαθηματική ικανότητα, δημιουργούμε μία ψυχομετρική κλίμακα αξιολόγησης, η οποία μπορεί να περιλαμβάνει 10 δοκιμασίες μαθηματικού τύπου, προοδευτικά δυσκολότερες. Για κάθε μία από τις δοκιμασίες αυτές, με βάση θεωρητικά, ερευνητικά, ή άλλα κριτήρια, βαθμολογούμε χρησιμοποιώντας κλίμακα μέτρησης ίσων διαστημάτων με

minimum 0 και *maximum* 10. Έτσι, η τελική συνολική βαθμολογία κάποιου που προσπάθησε να λύσει τα 10 αυτά μαθηματικά προβλήματα θα κυμαίνεται ανάμεσα στο 0 και στο 100. Η τελική όμως αυτή βαθμολογία, αν και φαίνεται να περιγράφει τη μαθηματική ικανότητα κάποιου με ποσοτικό τρόπο, στην ουσία αναφέρεται στην ποιοτική διαφοροποίηση του ατόμου αυτού από άλλα άτομα, και ενώ η ποιοτική αυτή διαβάθμιση είναι «αριθμοποιημένη» θα αποκτήσει ιδιαίτερο νόημα **μόνο** όταν μετατραπεί σε τυπικούς βαθμούς κατά τη διαδικασία της στάθμισης, βαθμούς δηλαδή που εκφράζουν τη σχετικότητα των αριθμών-βαθμολογιών του κάθε ατόμου ως προς τον γενικό πληθυσμό και ταυτόχρονα ως προς τη θεωρητική κανονική κατανομή του *Gauss*.

Η θεωρία των αληθών τιμών

Καθώς ο βασικός στόχος της ψυχομετρίας παραμένει η, όσο το δυνατόν, ακριβής μέτρηση των ανθρώπινων ιδιοτήτων, τίθεται το ερώτημα σχετικά με το κατά πόσο οι κλίμακες μέτρησης που χρησιμοποιούνται για τον καθορισμό των αρχικών τιμών ενός ατόμου σε κάποιο τεστ μπορούν να αποδώσουν την ακρίβεια αυτή. Το ίδιο ερώτημα τίθεται και για τις υποκειμενικές και ποιοτικού χαρακτήρα ψυχομετρικές κλίμακες αξιολόγησης.

Έχουμε ήδη αναφερθεί στην παράμετρο λάθους ή θορύβου, την παράμετρο e , η οποία και αντιπροσωπεύει το συστηματικό σφάλμα, την τυχειότητα και τις ατομικές διαφορές. Χρησιμοποιώντας αυτή την παράμετρο, η θεωρία των αληθών τιμών (*theory of true scores*) προσπαθεί να εκτιμήσει τις πραγματικές τιμές μιας μεταβλητής κατά την ψυχομετρική διαδικασία. Αν υποθέσουμε ότι η φαινομενική αρχική τιμή ενός ατόμου σε κάποια μεταβλητή είναι ίση με X , τότε $X = T + e$, όπου T είναι η αληθής τιμή και e είναι το σφάλμα της μέτρησης. Έστω, ότι κάποιος έχει αρχική τιμή $X = 103$. Αυτή η φαινομενική αρχική τιμή είναι δυνατό να προέρχεται από $T = 100$ και $e = 3$. Στην περίπτωση αυτή, η αληθής τιμή είναι 100. Κάλλιιστα όμως, θα μπορούσε η ίδια φαινομενική αρχική τιμή 103 να προέρχεται από $T = 50$ και $e = 53$. Στην δεύτερη αυτή περίπτωση, η πραγματική-αληθής τιμή του ατόμου ως προς την μετρώμενη μεταβλητή είναι 50. Από το παράδειγμα αυτό φαίνεται το πόσο δύσκολο είναι να εκτιμήσουμε την πραγματική τιμή για μια μέτρηση στην οποία έχουμε ήδη προβεί αν δεν γνωρίζουμε την παράμετρο σφάλματος.

Τρία είναι τα θεώρηματα που διέπουν και καθορίζουν τη θεωρία των αληθών τιμών: **α)** Η παράμετρος e κατανέμεται τυχαία στον γενικό πληθυσμό, δεν υπάρχει δηλαδή συστηματικό λάθος ή συστηματικός θόρυβος, **β)** οι αληθείς τιμές των μετρήσεων δεν παρουσιάζουν καμία συνάφεια με την παράμετρο λάθους και, **γ)** οι αληθείς τιμές που αναφέρονται σε διαφορετικές μεταβλητές, όπως αυτές μετρώνται για ένα άτομο, δεν αλληλεπιδρούν.

Όσον αφορά στο πρώτο θεώρημα, τα πράγματα είναι σχετικά απλά: η ίδια η θεωρητική κανονική κατανομή του *Gauss* προέρχεται από το θεώρημα αυτό και είναι δυνατό να υποθέσουμε ότι η συνθήκη του θεωρήματος θα ισχύει και για τον γενικό πληθυσμό λόγω της αλληλεξουδετέρωσης των αποκλίσεων σε ένα μεγάλο αριθμό ατόμων. Η λογική αυτή της αλληλοεξουδετέρωσης ισχύει και για τα σφάλματα που είναι ενσωματωμένα στις αποκλίσεις από τη μέση τιμή, επομένως το σφάλμα της μέτρησης είναι δυνατό να θεωρηθεί ότι κατανέμεται ομαλά στον γενικό πληθυσμό. Το δεύτερο θεώρημα είναι περισσότερο προβληματικό γιατί, συνήθως, υπάρχει συνάφεια μεταξύ των αληθών τιμών και των παραμέτρων σφάλματος (η πιο συνήθης περίπτωση είναι αυτή στην οποία μεγάλες αληθείς τιμές έχουν αντίστοιχα και μεγάλο δείκτη σφάλματος, ενώ οι μικρές τιμές έχουν μικρό δείκτη σφάλματος). Το φαινόμενο ορίζεται στη Στατιστική ως *ετεροσκεδασμός*, αναφέρεται σε ανομοιογενείς ως προς τη διασπορά πληθυσμούς και αντιμετωπίζεται συνήθως με στατιστικές και αλγεβρικές μεθόδους. Τέλος, σχετικά με το τρίτο θεώρημα, είναι πολύ σπάνια η περίπτωση κατά την οποία δεν παρατηρούνται αλληλεπιδράσεις δύο ή περισσότερων μεταβλητών που μετρώνται για το ίδιο άτομο. Είναι λοιπόν επόμενο, ότι, εφόσον τα θεώρημα αυτά (που αποτελούν παράλληλα και στατιστικές παραδοχές) δεν ισχύουν πλήρως, η εκτίμησή μας για την αληθή τιμή μέτρησης και για την αντίστοιχη παράμετρο λάθους δεν θα είναι επακριβής, παρά μόνο θα προσεγγίζει την πραγματικότητα. Από την καθαρά στατιστική άποψη, η αληθής τιμή μέτρησης είναι εκείνη που θα προέκυπτε εάν μετρούσαμε την ίδια μεταβλητή άπειρες φορές στο ίδιο άτομο. Αυτό είναι, βέβαια, πρακτικά ανέφικτο, όμως όσο θα προχωρούσαμε σε αυτή την επ' άπειρον διαδικασία, τόσο οι παράμετροι λάθους, λόγω ακριβώς της τυχαίας κατανομής τους, θα αλληλεξουδετερώνονταν και έτσι θα είχαμε μηδενικό λάθος. Στη λογική αυτή βασίζεται και η στατιστική πρακτική της εκτιμητικής διαδικασίας που ακολουθούμε όταν θελήσουμε, κατά τον έλεγχο ενός ψυχομετρικού εργαλείου, να

αποφανθούμε για το μέγεθος της ακρίβειας μέτρησης που διαθέτει το εργαλείο αυτό. Η περιγραφή της διαδικασίας αυτής θα γίνει σε επόμενα κεφάλαια.

Μέρος 3. Αξιοπιστία μετρήσεων και αξιοπιστία των τεστ

Γενικές έννοιες

Στην προσπάθειά μας να μετρήσουμε, με όσο το δυνατόν μεγαλύτερη ακρίβεια, τις διάφορες μεταβλητές στις οποίες αναφέρεται το εκάστοτε ψυχομετρικό εργαλείο, πέρα από το ενδιαφέρον μας για τα πιθανά σφάλματα μέτρησης και τα όσα συνοδεύουν την εκτίμηση των αληθών τιμών, γνωρίζουμε ότι η αξία μιας οποιασδήποτε αληθούς τιμής μέτρησης έγκειται και στο κατά πόσο είναι δυνατόν να θεωρήσουμε την τιμή αυτή σταθερή. Εάν μια ζυγαριά, μέσα σε διάστημα 5 λεπτών, μετρήσει το βάρος ενός ατόμου 10 φορές και προκύψουν 10 διαφορετικές ενδείξεις, τότε οι τιμές αυτές δεν είναι σταθερές και επομένως μας είναι άχρηστες διότι δεν γνωρίζουμε ποια από όλες είναι τελικά η πραγματική. Τα πράγματα γίνονται πολύ δύσκολα, όταν οι μετρήσεις δεν είναι φυσικά μεγέθη αλλά «ψυχοσωματικά χαρακτηριστικά του ατόμου»⁴, όταν δηλαδή, λόγω των λαθών μέτρησης, στα οποία περιλαμβάνεται πλέον και η ίδια η υφή του ψυχομετρικού εργαλείου καθώς δεν αφορά σε φυσικά-άμεσα μετρούμενα μεγέθη, έχουμε αστάθεια στις μετρήσεις αυτές.

Εξαιτίας της πιθανής αυτής αστάθειας, τίθεται το ζήτημα της αξιοπιστίας, είτε αυτή αναφέρεται στις μετρήσεις που προκύπτουν από μια ερευνητική διαδικασία ή ένα ψυχομετρικό εργαλείο είτε αναφέρεται γενικά σε ένα τεστ. Ως προς την αξιοπιστία των μετρήσεων, μας ενδιαφέρει το κατά πόσο μετρώντας την ίδια μεταβλητή *n* φορές κάτω από παρόμοιες συνθήκες, σε διαφορετικά χρονικά σημεία και σε παρόμοιους πληθυσμούς, θα λάβουμε παρόμοιες μετρήσεις. Ακόμη περισσότερο, όταν αναφερόμαστε στην αξιοπιστία ενός τεστ, μας ενδιαφέρει οι παρόμοιες μετρήσεις να προέρχονται από το ίδιο άτομο. Για παράδειγμα, χορηγώντας σε κάποιον ένα τεστ προσωπικότητας που μετρά την ψυχοπαθητική εκτροπή και λαμβάνοντας κάποια τιμή, θα πρέπει σε επόμενη χορήγηση του ίδιου

⁴ Βλέπε σύγγραμμα: **I.N. Παρασκευόπουλος** (1992) *Ψυχολογία Ατομικών Διαφορών*, σελ. 20.

τεστ, κάτω από παρόμοιες συνθήκες, να λάβουμε μια παρόμοια τιμή ψυχοπαθητικής εκτροπής. Αν δεν συμβεί αυτό, τότε ποια από τις δύο αυτές τιμές μπορούμε να εμπιστευτούμε; Επομένως, εάν δεν μπορούμε να έχουμε εμπιστοσύνη στις μετρήσεις που προκύπτουν από ένα τεστ, τότε το τεστ αυτό είναι αναξιόπιστο.

Φυσικά, υπάρχουν διάφορα είδη αξιοπιστίας τα οποία μας ενδιαφέρουν, ανάλογα με το είδος του τεστ, τις υπό μέτρηση μεταβλητές, τον πληθυσμό στον οποίο αναφέρονται οι μετρήσεις, τη δομή των τμημάτων-υποκλιμάκων του τεστ. Κατά βάση όμως, σε όλες τις περιπτώσεις, μας ενδιαφέρει να είμαστε βέβαιοι για το πόσο μπορούμε να βασιζόμαστε στις μετρήσεις που μας παρέχει ένα τεστ. Ας μην λησμονούμε ότι σε πάρα πολλές περιπτώσεις, από τις μετρήσεις αυτές και τα συμπεράσματα στα οποία θα καταλήξουμε, είναι πιθανό να εξαρτάται –έστω εν μέρει– η μελλοντική εξέλιξη ενός ατόμου, οι σχέσεις του με τους άλλους και κάθε πλευρά της καθημερινής του ζωής, ειδικά όταν το άτομο αυτό είναι παιδί. Είναι, λοιπόν, πολύ επικίνδυνο να αποφαινόμεστε για τα ψυχοσωματικά χαρακτηριστικά κάποιου, αν δεν γνωρίζουμε την αξιοπιστία του ψυχομετρικού εργαλείου που χρησιμοποιήσαμε για να τα ανακαλύψουμε.

Αξιοπιστία και σφάλματα μέτρησης

Στο κεφάλαιο αυτό θα εξετάσουμε κάποιες από τις ειδικές εκείνες παραμέτρους που προκαλούν σφάλματα μέτρησης, τα οποία έχουν επίπτωση στην αξιοπιστία των μετρήσεων, ή στην αξιοπιστία ενός τεστ γενικότερα.

Γνωρίζοντας ήδη την ύπαρξη της παραμέτρου e και την ιδιαίτερη σημασία της για την ψυχομετρία, θα πρέπει να ασχοληθούμε με το αν η παράμετρος αυτή επιτείνεται λόγω συγκεκριμένων παραγόντων και συνθηκών κατά τη διαδικασία της χορήγησης ενός τεστ ή και λόγω γενικότερων κατασκευαστικών προβλημάτων του τεστ. Ο θόρυβος (*noise*) που μπορεί να προκληθεί από τέτοιους παράγοντες, είναι δυνατό να διογκώσει πολύ την παράμετρο e , αλλά το σοβαρότερο πρόβλημα είναι ότι μπορεί να τη διογκώσει *ανεξέλεγκτα*. Η συνέπεια είναι ότι, μετά το πέρας των μετρήσεων, δεν γνωρίζουμε τις αληθείς τιμές που αφορούν στις μετρήσεις αυτές, καθώς δεν γνωρίζουμε επακριβώς την παράμετρο λάθους και ούτε μπορούμε να είμαστε βέβαιοι για τη σταθερότητα της παραμέτρου αυτής, άρα και των αληθών τιμών που τη συνοδεύουν. Σε αντιδιαστολή,

όταν μετρούμε ένα φυσικό μέγεθος (π.χ. μήκος δωματίου) σπάνια θα έχουμε μεγάλη και ανεξέλεγκτη παράμετρο σφάλματος (π.χ., μετρήσεις εκτός ευθείας), αλλιώς, σταθερά και αξιόπιστα, θα βρίσκουμε την αληθή τιμή μήκους του δωματίου, όσες φορές και αν μετρήσουμε την ίδια απόσταση. Στην περίπτωση, οι ανεξέλεγκτοι παράγοντες θορύβου είναι ιδιαίτερα περιορισμένοι ή ανύπαρκτοι, αλλά όταν οι μετρήσεις μας αφορούν στην προσωπικότητα, τη νοημοσύνη, τις κοινωνικές παραμέτρους (στάσεις, αξίες, ενδιαφέροντα, κοινωνικές δεξιότητες, κ.λπ.), τις ικανότητες κ.ο.κ., τότε είναι φυσικό επόμενο οι παράγοντες θορύβου να λειτουργούν στο μέγιστο βαθμό. Οι συνθήκες εκείνες που μπορούν να επιδράσουν αρνητικά και να αυξήσουν δραματικά και ανεξέλεγκτα την παράμετρο λάθους αναλύονται αμέσως παρακάτω:

α) *Η ίδια η διαδικασία χορήγησης του τεστ.* Όταν προσπαθούμε να μετρήσουμε, για παράδειγμα, τη νοημοσύνη ενός παιδιού, πρέπει να εφαρμόζουμε πιστά τους κανόνες χορήγησης που μας διαθέτει ο οδηγός εξεταστή, το εγχειρίδιο δηλαδή του τεστ. Οι συνήθειες κανόνες είναι ότι η εξέταση πρέπει να γίνει σε περιβάλλον ήσυχο, αρκετά απομονωμένο από θορύβους και άλλες πηγές παρενόχλησης, χωρίς την παρουσία τρίτων, ενώ ιδιαίτερης σημασίας είναι η ανάπτυξη σχέσεων εμπιστοσύνης με τον εξεταζόμενο. Τα διάφορα τεστ νοημοσύνης συνήθως έχουν ειδικούς κανόνες για κάθε τμήμα χορήγησής τους, οι οποίοι και πρέπει να τηρούνται πιστά (για παράδειγμα, αν ο χρόνος εξέτασης μιας συγκεκριμένης κλίμακας αφήνει περιθώριο 5 λεπτών στον εξεταζόμενο για να απαντήσει, το χρονικό αυτό όριο δεν πρέπει να ξεπεραστεί, δεν πρέπει δηλαδή ο εξεταστής να «κάνει χάρη» στον εξεταζόμενο. Στις περισσότερες μάλιστα περιπτώσεις, υπάρχουν όρια βάσης και οροφής⁵ τα οποία δεν πρέπει, επίσης, να

⁵ Τα όρια βάσης και οροφής, έχουν καθοριστεί για κάθε τεστ, ανάλογα με το είδος του τεστ, το είδος της υπό μέτρησης μεταβλητής και τα αποτελέσματα που προέκυψαν κατά τη στάθμιση του τεστ. Χωρίς, επί του παρόντος, να εισέλθουμε στη διαδικασία της στάθμισης, θα πρέπει να αναλογισθούμε ότι ένα παιδί ηλικίας 10 ετών μπορεί να απαντήσει με ευκολία στις πρώτες ερωτήσεις μιας κλίμακας νοημοσύνης. Γνωρίζοντας τον ακριβή αριθμό των ερωτήσεων αυτών, αρχίζουμε τη χορήγηση της κλίμακας από το αμέσως επόμενο ερώτημα, θεωρώντας δεδομένο ότι το παιδί έχει απαντήσει σωστά στις προηγούμενες ερωτήσεις και θέτοντας ειδικά κριτήρια για την περίπτωση που το παιδί αυτό δεν έχει τελικά την αναμενόμενη αυτή ικανότητα. Η όλη διαδικασία αναφέρεται στο φαινόμενο βάσης (*floor effect*) και περιγράφεται στα διάφορα τεστ με την ορολογία *basal*. Από την άλλη, υπάρχει και η επίδραση οροφής (*ceiling effect*) η οποία αναφέρεται στην ακριβώς αντίστροφη λογική. Ένα παιδί 10 ετών, σε σχέση πάντα με παιδιά μικρότερης και μεγαλύτερης ηλικίας, μπορεί να καλύψει ένα τμήμα της κλίμακας μέτρησης δίνοντας σωστές απαντήσεις, αλλά συνήθως, λόγω της προοδευτικής αύξησης του βαθμού δυσκολίας των ερωτημάτων, δεν μπορεί να απαντήσει σωστά σε ολόκληρη την κλίμακα αξιολόγησης. Γνωρίζοντας, και πάλι σύμφωνα με τα στοιχεία που έχουν προκύψει από τη στάθμιση, τα όρια επίδοσης που ισχύουν για τις διάφορες ηλικίες στις οποίες αναφέρεται το τεστ, καθορίζουμε τον αριθμό των ερωτημάτων που πρέπει να έχουν συνεχόμενα απαντηθεί εσφαλμένα για να διακόψουμε τη χορήγηση. Γνωρίζουμε έτσι την οροφή (*ceiling*) και, όταν ο εξεταζόμενος, λόγω συνεχών αποτυχιών, φανεί ότι δεν μπορεί να δώσει περαιτέρω σωστές απαντήσεις, διακόπτουμε την εξέταση. Τα φαινόμενα βάσης και οροφής καθώς και

παραβιάζονται. Όταν ο εξεταζόμενος αποτύχει σε τρία, για παράδειγμα, συνεχή ερωτήματα, τότε διακόπτουμε τη χορήγηση του τμήματος του τεστ και προχωρούμε στο επόμενο τμήμα. Δεν «δίνουμε άλλη μια ευκαιρία» στον εξεταζόμενο, διότι αυτή μας η κίνηση είναι δυνατόν να επηρεάσει την τελική μέτρηση και μάλιστα ανεξέλεγκτα. Θα πρέπει, επομένως, να γνωρίζουμε, εκ των προτέρων, πολύ καλά τη διαδικασία χορήγησης (ειδικά των τεστ νοημοσύνης) και να την εφαρμόζουμε πιστά, καθώς οι αποφάσεις που θα προκύψουν από τέτοιες μετρήσεις είναι ιδιαίτερα σημαντικές για τον εξεταζόμενο. Τέλος, η διαδικασία χορήγησης ενός τεστ θα πρέπει να είναι η ίδια για όλους. Εάν χορηγούμε ένα τεστ οποιουδήποτε τύπου και αντικειμένου σε κάποιον με έναν τρόπο και το ίδιο τεστ σε άλλον με άλλον τρόπο, δεν υπάρχει πλέον δυνατότητα σύγκρισης των δύο αποτελεσμάτων-μετρήσεων. Καθώς όμως η συγκρισιμότητα αυτή είναι ιδιαίτερα σημαντικό χαρακτηριστικό της ψυχομετρικής διαδικασίας, παραβιάζοντάς την, παραβιάζουμε και ολόκληρη την ψυχομετρική βάση της συγκεκριμένης μέτρησης και δεν μπορούμε να καταλήξουμε σε συμπεράσματα.

β) Η επίδραση τυχαίων γεγονότων. Είναι προφανές ότι, κατά τη χορήγηση ενός τεστ, οποιαδήποτε διακοπή επιφέρει αλλοίωση των αποτελεσμάτων, με τρόπο που δεν είναι δυνατό να προβλεφθεί. Τα τυχαία αυτά γεγονότα δεν συμβαίνουν απαραίτητα κατά τη διαδικασία της εξέτασης, αλλά μπορεί να έχουν συμβεί και πριν από αυτήν. Για παράδειγμα, κατά τη διαδικασία της εξέτασης, υπάρχει περίπτωση να μπει στο δωμάτιο κάποιος για οποιονδήποτε λόγο και να διασπάσει την προσοχή του εξεταζομένου. Επίσης, είναι πιθανόν, όταν η εξέταση γίνεται για πολλούς εξεταζομένους ταυτόχρονα (π.χ. ένα τεστ αντιληπτικής ικανότητας και ταχύτητας με σκοπό την πλήρωση θέσεων σε μία εταιρεία), κάποιος από αυτούς να παραβιάσει τους κανόνες της εξέτασης (π.χ. να συνεχίσει να δίνει απαντήσεις και μετά τη λήξη του χρόνου). Αυτή η παραβίαση και η παρατήρηση που θα γίνει από τον εξεταστή, μπορούν να επηρεάσουν τις απαντήσεις των άλλων εξεταζομένων. Πέρα όμως από τα τυχαία, τέτοιου τύπου, γεγονότα, υπάρχουν και οι περιπτώσεις που ο εξεταζόμενος προσέρχεται στην εξέταση επηρεασμένος από κάτι που έχει συμβεί αμέσως πριν. Για παράδειγμα μία έντονη διαμάχη μεταξύ των παιδιών στο διάλειμμα του σχολείου θα προκαλέσει συναισθηματικές αντιδράσεις και θα μας φέρει στον χώρο εξέτασης ένα παιδί νευρικό,

εφαρμογής τους στην εξεταστική διαδικασία θα αναλυθούν περαιτέρω στα Κεφάλαια που αφορούν στη στάθμιση των τεστ.

σε υπερένταση, με υπερκινητικές τάσεις και κάθε άλλο παρά συγκεντρωμένο στο αντικείμενο της εξέτασης αυτής. Συνοπτικά, θα πρέπει να προσπαθούμε να διενεργούμε την οποιαδήποτε εξέταση αποφεύγοντας την επίδραση οποιουδήποτε εξω-εξεταστικού παράγοντα. Βέβαια, εάν ένα τυχαίο γεγονός είναι αναπόφευκτο και, όποια προφύλαξη και να πάρουμε, τελικά συμβεί (π.χ. ένας σεισμός), φυσικά δεν θα συνεχίσουμε τη χορήγηση, διότι δεν θα μπορούμε σε καμία περίπτωση να εμπιστευθούμε το οποιοδήποτε αποτέλεσμα.

γ) *Η διάθεση του εξεταζομένου.* Για οποιαδήποτε εξέταση, σε όποια μεταβλητή και αν αναφέρεται, η προσωπική διάθεση του εξεταζομένου θα πρέπει να είναι όσο το δυνατόν ουδέτερη. Εάν ο εξεταζόμενος είναι φορτισμένος θετικά ή αρνητικά απέναντι στο τεστ, εάν έχει άγχος, εάν αισθάνεται «άβολα» ή εάν είναι εκνευρισμένος για οποιονδήποτε λόγο, θα πρέπει πριν χορηγήσουμε το τεστ να οικοδομήσουμε σχέσεις εμπιστοσύνης μαζί του. Αυτή είναι μια δύσκολη διαδικασία και πολλές φορές δεν επιτυγχάνεται· στις περιπτώσεις αυτές, είναι προτιμότερο να μην χορηγήσουμε το τεστ και να επανέλθουμε σε άλλη χρονική στιγμή. Ο τρόπος με τον οποίο διάκινται οι εξεταζόμενοι απέναντι σε ένα ή σε όλα τα τεστ γενικά και, οι οποιεσδήποτε φορτίσεις, θετικές ή αρνητικές, συνοδεύουν την εξέταση, ονομάζονται *επιδράσεις συμμετεχόντων* και αποτελούν πολύ επικίνδυνο παράγοντα για την ανεξέλεγκτη διόγκωση της παραμέτρου σφάλματος και επομένως συνιστούν άμεσο κίνδυνο για την αξιοπιστία των αποτελεσμάτων.

δ) *Οι επιδράσεις του εξεταστή.* Εκτός από τις επιδράσεις των συμμετεχόντων, υπάρχει και ένας άλλος πολύ σοβαρός κίνδυνος για την αξιοπιστία των αποτελεσμάτων, ο οποίος και αποτελεί, μάλιστα, λανθάνοντα παράγοντα της συμπεριφοράς και υπεισέρχεται στις λεκτικές και μη-λεκτικές αντιδράσεις του εξεταστή, πριν και κατά τη διάρκεια της χορήγησης του τεστ. Οι μελέτες που έχουν γίνει στον χώρο της διερεύνησης και αποφυγής των φαινομένων της επίδρασης του ερευνητή-εξεταστή είναι πολυάριθμες, αλλά αρκεί να αναλογισθούμε το παράδειγμα του *clever Hans*, του αλόγου που ήξερε να δίνει τη σωστή απάντηση σε αριθμητικές πράξεις χτυπώντας το πόδι στο έδαφος και μετρώντας τη σωστή απάντηση.⁶ Ο κυριότερος ερευνητής και

⁶ Ο *clever Hans*, με τις αριθμητικές του ικανότητες, είχε γίνει το επίκεντρο του ενδιαφέροντος στις αρχές του αιώνα. Προσεκτική παρατήρηση, όμως, από τον *Pfungst* οδήγησε στην ανακάλυψη ότι ο 'εκπαιδευτής' του αλόγου *Von Osten* είχε μάθει στο άλογο, μέσω συντελεστικής μάθησης, να χτυπά το πόδι στο έδαφος όταν το

θεωρητικός στον χώρο των επιδράσεων ερευνητή είναι ο *Rosenthal*, ο οποίος για σειρά ετών διερεύνησε και ανέλυσε τις επιδράσεις αυτές και τον τρόπο λειτουργίας τους σε διάφορες πειραματικές συνθήκες.

Η σημαντικότερη ίσως επίδραση ερευνητή είναι η «χωρίς πρόθεση υποβολή του αναμενόμενου αποτελέσματος» (*mediation of expectancy*) από τον ερευνητή-εξεταστή στον εξεταζόμενο. Η *Anastasi* αναφέρεται σε ένα χαρακτηριστικό παράδειγμα (*Masling*, 1965) τέτοιας υποβολής: Χρησιμοποιώντας το προβολικό τεστ προσωπικότητας *Rorschach*, 14 μεταπτυχιακοί φοιτητές εξέτασαν 14 δείγματα συμμετεχόντων με σκοπό την καταγραφή του είδους των απαντήσεων των συμμετεχόντων αυτών. Οι 14 λοιπόν μεταπτυχιακοί φοιτητές λειτούργησαν ως εξεταστές, μόνο που οι 7 από αυτούς είχαν πληροφορηθεί σε ειδικό σεμινάριο ότι στο *Rorschach* οι συμμετέχοντες γενικά βλέπουν μορφές ζώων σε πολύ μεγαλύτερο βαθμό από ό,τι μορφές ανθρώπων. Οι άλλοι 7 «εξεταστές», και πάλι σε ειδικό σεμινάριο, έντεχνα είχαν πληροφορηθεί ότι οι συμμετέχοντες βλέπουν πολύ περισσότερο μορφές ανθρώπων παρά ζώων. Στη διαδικασία της εξέτασης των συμμετεχόντων από τους 14 εξεταστές, όπως κατόπιν φάνηκε από τις μαγνητοφωνημένες συνομιλίες τους, δεν υπήρξε καμία λεκτική υποβολή του αναμενόμενου αποτελέσματος. Κανείς επίσης από τους 14 μεταπτυχιακούς φοιτητές ούτε από τους συμμετέχοντες δεν αντιλήφθηκε την οποιαδήποτε χωρίς πρόθεση υποβολή κάποιου αναμενόμενου αποτελέσματος. Ενώ δεν υπήρχε, από ό,τι φαινόταν, καμία επίδραση ερευνητή τέτοιου τύπου, οι δύο ομάδες των 7 εξεταστών είχαν τελικά καταγράψει, σε στατιστικά σημαντικά διαφορετικό ποσοστό, απαντήσεις των συμμετεχόντων που επιβεβαίωναν τις πληροφορίες που η κάθε ομάδα είχε εκ των προτέρων για την αναλογία ζώων-ανθρώπων στις μορφές που έβλεπαν οι συμμετέχοντες στο πείραμα. Παρόμοια αποτελέσματα προέκυψαν από έρευνες του *Rosenthal* αλλά και άλλων ερευνητών και οδήγησαν στη διατύπωση της θεωρίας της «αυτοεκπληρούμενης προφητείας» (*self-fulfilling prophecy*), η οποία και αποτελεί το

ερέθισμα είχε δεδομένη μορφή (τόνο φωνής, στάση σώματος του εκπαιδευτή, ακροατήριο σε αναμονή της απάντησης του *Hans*, κ.τ.λ.) και να δίνει το σωστό αποτέλεσμα σε μια αριθμητική πράξη, έστω 3 επί 24 διά 6. Στην περίπτωση αυτή ο *Hans* θα απαντούσε '12', χτυπώντας διαδοχικά το πόδι στο έδαφος 12 φορές. Ο *Von Osten*, κατά την 'ιεροτελεστία' αυτή είχε τα μάτια χαμηλωμένα στο έδαφος και, όταν ο *Hans* πλησίαζε στο σωστό αποτέλεσμα, σήκωνε τα μάτια κοιτάζοντας το άλογο. Το ερέθισμα αυτό διέκοπτε τη διαδικασία και ο *Hans* έβρισκε τη σωστή απάντηση. Το πιο σημαντικό, βέβαια, είναι ότι, όπως αναλύει ο *Pfungst*, ο εκπαιδευτής *Von Osten* δεν είχε αντιληφθεί ότι με τον τρόπο αυτόν επηρέαζε το άλογο και έτσι, χωρίς και ο ίδιος να το συνειδητοποιεί, επιδρούσε στο αποτέλεσμα, πιστεύοντας και αυτός ότι ο *Hans* ήταν πράγματι 'clever'.

κύριο μηχανισμό πρόκλησης της χωρίς πρόθεση υποβολής του αναμενόμενου αποτελέσματος.⁷

Οι επιδράσεις του ερευνητή-εξεταστή στη διαδικασία και, τελικά, στο αποτέλεσμα που προκύπτει με βάση ένα ψυχομετρικό εργαλείο, δεν θα πρέπει σε καμία περίπτωση να θεωρούνται αμελητέες. Στον Πίνακα που ακολουθεί παρουσιάζονται τα συνοπτικά αποτελέσματα των *Rosenthal & Rubin* (1978) για 345 ερευνητικές μελέτες σε διάφορους θεωρητικούς τομείς της ψυχολογίας, στις οποίες μελέτες εξετάστηκε το αν υπήρξε επίδραση της χωρίς πρόθεση υποβολής του αναμενόμενου αποτελέσματος.

Ερευνητικός τομέας	Αριθμός μελετών	Ποσοστό μελετών όπου παρατηρήθηκε Χ.Π.Υ.Α.Α.
Χρόνος αντίδρασης (<i>reaction times</i>)	9	22%
Προβολικά τεστ κηλίδων (<i>inkblot tests</i>)	9	44%
Μάθηση στα ζώα (<i>animal learning</i>)	15	73%
Συνεντεύξεις σε εργαστήριο (<i>laboratory interviews</i>)	29	38%
Ψυχοφυσιολογικές μετρήσεις (<i>psychophysical judgments</i>)	23	43%
Μάθηση και Ικανότητες (<i>learning and ability</i>)	34	29%
Αντίληψη προσώπων (<i>person perception</i>)	119	27%
Καθημερινές καταστάσεις (<i>everyday situations</i>)	112	40%
Σύνολο μελετών	350*	Μέσο ποσοστό = 39,5%

* 5 μελέτες αναφέρονται ταυτόχρονα σε δύο ερευνητικούς τομείς.

[Πηγή: **Rosenthal, R. & Rubin, D.B.** (1978) *Interpersonal Expectancy Effects: the first 345 studies. The Behavioral and Brain Sciences*, 3, 317-415, στοιχεία του Πίνακα ς 1.]

Ανακεφαλαιώνοντας, οι παράγοντες που μπορεί να λειτουργούν και να επηρεάζουν τις μετρήσεις μιας ψυχομετρικής διαδικασίας είναι πολλοί και, όσο και να προσπαθήσει ο εξεταστής να τους αποφύγει, είναι βέβαιο ότι δεν θα το πετύχει απόλυτα. Επιπρόσθετα, θα πρέπει να αναφέρουμε και τις 5 πηγές θορύβου, όπως τις όρισε ο *Thorndike* το 1947.

1. *Διαρκής και γενική*: γενική ικανότητα στα τεστ.
2. *Διαρκής και ειδική*: γνώση ή άγνοια ως προς την απάντηση σε συγκεκριμένο ερώτημα.
3. *Προσωρινή και γενική*: αδιαφορία ή κόπωση σε σχέση με την συμμετοχή σε οποιοδήποτε τεστ.
4. *Προσωρινή και ειδική*: μια ανεπάρκεια (νοητική ή, άλλης μορφής) που εμποδίζει την επιτυχή ενασχόληση με κάποια είδη ερωτημάτων.
5. Επιτυχείς απαντήσεις λόγω τύχης.

Συμπερασματικά, οι ατομικές διαφορές και τα σφάλματα μέτρησης που προκαλούνται από τις διάφορες πηγές θορύβου, καταλήγουν σε διόγκωση της

⁷ Για περισσότερες πληροφορίες σχετικά με το φαινόμενο της αυτοεκπληρούμενης προφητείας, βλέπε το σύγγραμμα: **Δ. Γεώργας** (1995) *Κοινωνική Ψυχολογία*, Τόμος Α', σελ. 200 & 272 και Τόμος Β', σελ. 167-169.

παραμέτρου σφάλματος, η οποία και πρέπει να υπολογισθεί και να χρησιμοποιηθεί για την εκτίμηση των αληθών τιμών. Ο τρόπος για να επιτύχουμε στη διαδικασία αυτή είναι να υπολογίσουμε την αξιοπιστία των μετρήσεων που προκύπτουν από ένα ψυχομετρικό εργαλείο κατά τις φάσεις της προσαρμογής και στάθμισής του σε κάποιον συγκεκριμένο πληθυσμό.

Από τη στιγμή, βέβαια, που αναφερόμαστε σε πληθυσμό, είμαστε υποχρεωμένοι να αναφερθούμε πρώτα σε δείγμα, ή καλύτερα σε δείγματα και μάλιστα αυτά που αφορούν στη θεωρία των δειγματοληπτικών κατανομών.⁸ Ο υπολογισμός, άλλωστε, των δεικτών αξιοπιστίας βασίζεται στα θεωρήματα που προκύπτουν από τη θεωρία αυτή, η οποία ερμηνεύει και καθιστά χρήσιμους τους δείκτες αυτούς. Στη θεωρία των δειγματοληπτικών κατανομών, λόγω της εκτιμητικής μαθηματικής βάσης που τη στηρίζει, είμαστε σε θέση να προσεγγίσουμε όχι μόνο τις μετρήσεις που προκύπτουν μέσω ενός, υπό έλεγχο, ψυχομετρικού εργαλείου, αλλά και το μέγεθος σφάλματος των μετρήσεων αυτών. Είμαστε, δηλαδή, σε θέση να καθορίσουμε με αρκετά μεγάλη ακρίβεια την παράμετρο «λάθους». Έχοντας καθορίσει την παράμετρο αυτή, θα είμαστε και σε θέση να εκτιμήσουμε με μεγάλη ακρίβεια τις αληθείς τιμές των μετρήσεών μας. Το σφάλμα που προκύπτει από τη δειγματοληψία, και περιλαμβάνει όλες τις πηγές συστηματικού θορύβου και φυσικά τις ατομικές διαφορές, λέγεται τυπικό σφάλμα μέτρησης (*standard error of measurement*). Ο βασικός μας στόχος, είναι να περιορίσουμε τα σφάλματα που προέρχονται από τις πηγές θορύβου, κατά το μέγιστο δυνατόν, και να παραμείνει στο τυπικό σφάλμα μέτρησης μόνο το τμήμα που οφείλεται στις ατομικές διαφορές. Φυσικά, αυτό είναι πολύ δύσκολο, αλλά εξαρτάται από την αντιπροσωπευτικότητα του δείγματος, την καταλληλότητά του, το μέγεθός του, από τη μέθοδο τυχαίας δειγματοληψίας που έχουμε εφαρμόσει κατά τη διαδικασία των μετρήσεων, καθώς, βέβαια, και από τη φύση των μετρήσεων που έχουμε διενεργήσει. Ως προς το θέμα της αντιπροσωπευτικότητας του δείγματος, οι μέθοδοι τυχαίας δειγματοληψίας είναι συνυφασμένες με την έννοια αυτή⁹ η οποία, βέβαια, στηρίζεται και στην *καταλληλότητα* του δείγματός μας. Δεν είναι δυνατόν να προσπαθούμε να ελέγξουμε

⁸ Βλέπε σύγγραμμα: **I.N. Παρασκευόπουλος** (1990) *Στατιστική, εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμος Β', Επαγωγική Στατιστική, Κεφάλαιο 8.

ψυχομετρικά ένα τεστ προσωπικότητας που περιλαμβάνει ερωτήσεις διατυπωμένες για παιδιά ηλικίας έως 16 ετών και να προβαίνουμε σε δειγματοληψία από τον στατιστικό πληθυσμό των ενηλίκων. Στην καταλληλότητα και αντιπροσωπευτικότητα του δείγματος, περιλαμβάνεται και το μέγεθος του δείγματος. Ένας βασικός κανόνας είναι ότι, για να εφαρμόσουμε παραμετρικά κριτήρια, δηλαδή κριτήρια ελέγχου που στηρίζονται στις δειγματοληπτικές κατανομές και, για να μπορέσουμε να χρησιμοποιήσουμε τις εκτιμητικές διαδικασίες, θα πρέπει στο δείγμα μας να έχουμε περισσότερα από 30 συμμετέχοντες. 31 όμως συμμετέχοντες δεν εξασφαλίζουν την αντιπροσωπευτικότητα του δείγματος, ενώ ένας άλλος παράγοντας που καθορίζει το μέγεθος (N) του δείγματος είναι ο αριθμός των δοκιμασιών-ερωτημάτων που περιλαμβάνονται στην υπό έλεγχο ψυχομετρική κλίμακα αξιολόγησης. Για παράδειγμα, εάν μια ψυχομετρική κλίμακα αξιολόγησης **μιας** μόνο μεταβλητής περιλαμβάνει 150 δοκιμασίες που συνολικά καταλήγουν στη μέτρηση της μεταβλητής αυτής, θα πρέπει, κάτω από συγκεκριμένες συνθήκες, το N να είναι τουλάχιστον 600 συμμετέχοντες, δηλαδή 4 φορές μεγαλύτερο, για να επιτρέψει την ασφαλή στατιστική ανάλυση των μετρήσεων που προκύπτουν για τα 150 αυτά επιμέρους ερωτήματα. Τονίζουμε τη λέξη «μιας» μεταβλητής, καθώς μια ψυχομετρική κλίμακα αξιολόγησης μπορεί να έχει 566 ερωτήματα (*MMPI*) αλλά δεν αναφέρονται όλα σε μία μεταβλητή, ούτε καταλήγουν σε μία μέτρηση. Παρόμοια δομή έχει και το *WISC-III* αλλά περιλαμβάνει 13 επιμέρους ψυχομετρικές υποκλίμακες αξιολόγησης. Εάν όλα τα ερεθίσματα του *MMPI* κατέληγαν σε μία μέτρηση, τότε, για να μπορέσουμε να ελέγξουμε την αξιοπιστία και την εσωτερική δομή των μετρήσεων αυτών, θα έπρεπε να συλλέξουμε δείγμα περίπου 2000 ατόμων.

Θα πρέπει να αναφερθεί ότι η αξιοπιστία των μετρήσεων αυξάνεται δραματικά όσο αυξάνουμε το N το δείγματος καθώς είναι γνωστό πως 100 άτομα στο δείγμα μας αντιπροσωπεύουν περίπου 150 άτομα στο στατιστικό πληθυσμό ενώ 300 άτομα στο δείγμα μας αντιπροσωπεύουν 1.150 άτομα στο στατιστικό πληθυσμό. Από τα 350 άτομα δείγματος και επάνω, η αντιπροσώπευση ατόμων του στατιστικού πληθυσμού αυξάνεται ραγδαία, για παράδειγμα, όταν $N_{\text{δείγματος}} = 350$, τότε $N_{\text{πληθυσμού που αντιπροσωπεύεται}} = 1.700$ άτομα και όταν το $N_{\text{δείγματος}} = 380$, τότε $N_{\text{πληθυσμού που αντιπροσωπεύεται}} = 1.000.000$

⁹ Περισσότερες πληροφορίες για τις μεθόδους τυχαίας δειγματοληψίας μπορούν να αναζητηθούν στο

άτομα. Αυτή η σχέση, δεν εξασφαλίζει απόλυτα την αντιπροσωπευτικότητα του δείγματος, αλλά αναφέρεται στην αριθμητική αντιπροσώπευση του στατιστικού πληθυσμού μέσω του δείγματος αυτού, δηλαδή, ακόμη και εάν αριθμητικά έχουμε επιτύχει αντιπροσωπευτικότητα, δεν είναι απαραίτητο ότι έχουμε τηρήσει και τους κανόνες της τυχαίας δειγματοληψίας. Σύμφωνα με τον Παρασκευόπουλο (1993), εάν υπάρχει συστηματική πηγή σφάλματος στις μετρήσεις μας και δεν τη γνωρίζουμε, αυξάνοντας απλά το δείγμα δεν αποφεύγουμε τον συστηματικό αυτό θόρυβο αλλά και τον διογκώνουμε. Οι παρατηρήσεις αυτές αναφέρονται στο γενικότερο πρόβλημα ανομοιογένειας που είναι περισσότερο πιθανό να παρατηρηθεί σε ένα μεγάλο σε μέγεθος παρά σε ένα μικρού μεγέθους δείγμα. Από την άλλη μεριά, με ένα δείγμα 40 ατόμων είναι μάλλον απίθανο να επιτύχουμε υψηλούς δείκτες αξιοπιστίας, ενώ με ένα δείγμα 200 ατόμων έχουμε πολύ περισσότερες πιθανότητες. Τέλος, όπως θα συζητηθεί και πιο κάτω, όσο μεγαλύτερος είναι ο αριθμός των δοκιμασιών-ερωτημάτων μιας ψυχομετρικής κλίμακας αξιολόγησης τόσο η εσωτερική συνοχή της κλίμακας αυτής βελτιώνεται και, τουλάχιστον ως προς τον «δείκτη αξιοπιστίας εσωτερικής συνέπειας», θα έχουμε δραματική βελτίωση.

Στα κεφάλαια που ακολουθούν, θα προσπαθήσουμε, όσο το δυνατόν συνοπτικά, να παρουσιάσουμε τα βασικότερα σημεία ως προς τα είδη αξιοπιστίας και τη χρησιμότητα του κάθε είδους. Ο Πίνακας που ακολουθεί αντιστοιχεί τα διάφορα είδη αξιοπιστίας με το είδος του δειγματοληπτικού λάθους στο οποίο αναφέρονται.

Δείκτης αξιοπιστίας (<i>reliability coefficient</i>)	Διασπορά λάθους
Επαναληπτικών μετρήσεων (<i>test-retest</i>)	λόγω διαφορετικών χρονικών σημείων
Εναλλακτικών τύπων (<i>alternate forms</i>)	λόγω διαφορετικού περιεχομένου
Ημικλάστων (<i>split-half</i>)	λόγω διαφορετικού περιεχομένου
Εσωτερικής συνέπειας (<i>internal consistency</i>)	λόγω διαφορετικού περιεχομένου και ετερογένειας περιεχομένου
Μεταξύ βαθμολογητών (<i>inter-scorer</i>)	λόγω διαφορών μεταξύ των βαθμολογητών

[Πηγή: **Anastasi, A.** (1990) *Psychological Testing*, έκτη έκδοση, N.Y., Macmillan Publishing Company, σελ. 126.]

Δείκτης αξιοπιστίας επαναληπτικών μετρήσεων

Εφόσον, για την αξιοπιστία ενός τεστ, μας ενδιαφέρει η σταθερότητα των αποτελεσμάτων, η καλύτερη απόδειξη της σταθερότητας αυτής είναι οι επαναληπτικές μετρήσεις που λαμβάνονται από δείγμα συμμετεχόντων. Εάν, ένα τεστ χορηγηθεί μία φορά (και για κάθε συμμετέχοντα του δείγματος καταλήξει σε μια μέτρηση) και, εάν υπάρξει συμφωνία της μέτρησης με μια δεύτερη μέτρηση του ίδιου συμμετέχοντα με το ίδιο τεστ (και αυτό προκύψει γενικά για το δείγμα), τότε το τεστ αυτό διαθέτει αξιοπιστία επαναληπτικών μετρήσεων. Είναι προφανές ότι οι επαναληπτικές αυτές μετρήσεις δεν είναι αναγκαία δύο, αλλά είναι δυνατό να είναι πολύ περισσότερες από δύο. Φυσικά, όσο περισσότερες είναι οι μετρήσεις αυτές, τόσο περισσότερες πληροφορίες έχουμε στα χέρια μας για την αξιοπιστία επαναληπτικών μετρήσεων του τεστ. Ο έλεγχος μέσω της αξιοπιστίας επαναληπτικών μετρήσεων στοχεύει να «θεραπεύσει» τον θόρυβο που προκύπτει σχετικά με τις μετρήσεις ενός τεστ, μόνο και μόνο εξαιτίας της χρονικής στιγμής που το τεστ αυτό χορηγείται. Προσπαθεί δηλαδή να εξαλείψει τον παράγοντα «χρόνο», προσπάθεια αρκετά δύσκολη, αλλά ευρέως επιθυμητή.

Η εκτίμηση του μεγέθους της αξιοπιστίας επαναληπτικών μετρήσεων γίνεται μέσω των δεικτών συνάφειας. Στις περισσότερες περιπτώσεις, καθώς τα περισσότερα τεστ καταλήγουν σε ασυνεχείς ή και συνεχείς αριθμητικές μετρήσεις, ο δείκτης συνάφειας που χρησιμοποιείται είναι ο *Pearson r*. Ο δείκτης αυτός θα πρέπει να είναι, βέβαια, θετικός (αν είναι αρνητικός, τότε μάλλον υπάρχει κάποιο πολύ σοβαρό λάθος στο τεστ) και, βέβαια, θα πρέπει να είναι ιδιαίτερα μεγάλου βαθμού, τουλάχιστον της τάξης του 0,70. Θα πρέπει να διευκρινίσουμε, όμως, ότι, πολλές φορές, λόγω της φύσης του ίδιου του τεστ, ο δείκτης συνάφειας μεταξύ των επαναληπτικών μετρήσεων μπορεί να είναι πολύ μικρότερος ως προς το βαθμό. Υπάρχουν, για παράδειγμα, τεστ τα οποία αναφέρονται σε έννοιες και μεταβλητές που είναι ευαίσθητες ακόμη και σε μικρή πάροδο του χρόνου. Η μέτρηση της ιδιοσυγκρασίας δεν είναι ευαίσθητη στην πάροδο αυτή, ή τουλάχιστον επηρεάζεται ελάχιστα, η μέτρηση, όμως, των κοινωνικών στάσεων και πεποιθήσεων είναι ιδιαίτερα ευάλωτη, λόγω της καθημερινής διαφοροποίησης που συντελείται στο άτομο ως προς τις μεταβλητές αυτές. Στην περίπτωση αυτή, υπάρχουν δύο τρόποι αντιμετώπισης: είτε κανείς αναζητεί-προσπαθεί να δημιουργήσει ένα τεστ το οποίο

θα είναι, όσο το δυνατόν, λιγότερο ευαίσθητο στις περιβαλλοντικές επιρροές και θα αναφέρεται κυρίως στις σταθερές παραμέτρους της αντίστοιχης συμπεριφοράς είτε χρησιμοποιεί για τον έλεγχο της αξιοπιστίας άλλες μεθόδους και όχι την αξιοπιστία επαναληπτικών μετρήσεων. Συχνά, πάντως, θα συναντήσουμε στην βιβλιογραφία αναφορές σε μικρού βαθμού δείκτες αξιοπιστίας επαναληπτικών μετρήσεων για διάφορα τεστ, κυρίως τεστ προβολικού τύπου. Αυτό δεν σημαίνει ότι τα τεστ αυτά είναι αναξιόπιστα, απλά είναι τα πιο αξιόπιστα από αυτά που ήδη υπάρχουν και μέχρι να δημιουργηθούν νέα, περισσότερο αξιόπιστα ως προς τις επαναληπτικές μετρήσεις τεστ, χρησιμοποιούμε τα υπάρχοντα.

Η μέτρηση της αξιοπιστίας επαναληπτικών μετρήσεων δεν είναι εύκολη υπόθεση. Παρουσιάζει πολλές πρακτικές δυσκολίες, με κυριότερη αυτή της συλλογής δεδομένων δεύτερης, τρίτης, κ.τ.λ. μέτρησης από το ίδιο δείγμα συμμετεχόντων. Εδώ υπάρχουν δύο εμπόδια: το *πρώτο* είναι η πρακτική δυσκολία του να βρούμε και να παρακολουθήσουμε ως προς τις μετρήσεις μία ομάδα ατόμων, για την οποία θα πρέπει να υπάρχουν τα κατάλληλα στοιχεία ταυτότητας για κάθε συμμετέχοντα, ώστε να μπορούμε να αντιστοιχίσουμε τις μετρήσεις τους κατά μήκος των επαναληπτικών χορηγήσεων του τεστ. Γνωρίζουμε, όμως, ότι τα περισσότερα τεστ, κατά τον δειγματοληπτικό τους έλεγχο, είναι ανώνυμα και αν δεν είναι, τότε, πολλές φορές οι συμμετέχοντες θα πολώσουν τις απαντήσεις τους εξαιτίας της έλλειψης ανωνυμίας. Θα πρέπει να βρούμε, λοιπόν, κάποιον άλλο τρόπο, ώστε να μπορέσουμε, εκ των υστέρων, να αντιστοιχίσουμε τις μετρήσεις του κάθε συμμετέχοντα (κωδικοί αριθμοί, ημερομηνίες γέννησης, κ.ά.). Ακόμη και αυτή η προσέγγιση όμως, μπορεί να δημιουργήσει πολώσεις και κάθε είδους πρακτικά προβλήματα. Για να προχωρήσουμε, επομένως, σε επαναληπτικές μετρήσεις, θα πρέπει πρώτα να έχουμε επιλύσει το πρακτικό αυτό πρόβλημα, άλλως θα καταλήξουμε με στοιχεία μη-εκμεταλλεύσιμα. Το *δεύτερο*, και πολύ σοβαρότερο, πρόβλημα με τις επαναληπτικές μετρήσεις είναι η επίδραση που είναι δυνατό να έχει η πρώτη μέτρηση στη δεύτερη. Χορηγώντας, για παράδειγμα, ένα τεστ νοημοσύνης σε κάποιο δείγμα, οι συμμετέχοντες του δείγματος έρχονται σε επαφή με το τεστ, καταλαβαίνουν τη λογική που το διέπει και καταχωρούν γνωστικές πληροφορίες ως προς το είδος, τη μορφή και το αντικείμενο των ερωτημάτων, κ.λπ. Σε μια δεύτερη μέτρηση, μετά από, έστω, 3 μήνες, είναι πολύ πιθανόν οι πληροφορίες αυτές να

ανασυρθούν από τη μακροπρόθεσμη μνήμη και να επιδράσουν στη νέα αυτή μέτρηση. Ακόμη και αν δεν υπάρξει συνειδητά τέτοια προσπάθεια εκ μέρους των συμμετεχόντων, πάντα θα υπάρχει μια υποσυνείδητη επίδραση που οφείλεται μόνο και μόνο στο γεγονός ότι το τεστ αυτό «το έχουν ξαναδεί». Στις περιπτώσεις αυτές, προσπαθούμε μεταξύ της πρώτης και δεύτερης μέτρησης να μεσολαβεί αρκετά μεγάλο χρονικό διάστημα, ώστε να απαλείψουμε όσο το δυνατόν τις επιδράσεις της προηγούμενης μέτρησης. Το διάστημα όμως αυτό, δεν θα πρέπει να είναι πολύ μεγάλο, γιατί σε διάστημα 2 ετών, για παράδειγμα, κάποιο τμήμα του δείγματος θα αναπτυχθεί νοητικά **με διαφορετικό τρόπο** από ό,τι ένα άλλο τμήμα του δείγματος και η εξέλιξη αυτή είναι, για την αξιοπιστία, καταστροφική.

Τα προβλήματα αυτά δεν ισχύουν, φυσικά, μόνο για τα τεστ νοημοσύνης, αλλά για κάθε ψυχομετρικό εργαλείο. Ο ερευνητής, στις περιπτώσεις αυτές, δεν έχει άλλη εναλλακτική παρά να αναζητήσει μια χρυσή τομή, ή να εφαρμόσει άλλες διαδικασίες για την ανεύρεση της αξιοπιστίας του τεστ.

Αξιοπιστία εναλλακτικών τύπων ή ισοδύναμων τύπων

Ένας άλλος τρόπος για να αποφύγουμε την επίδραση της πρώτης στη δεύτερη μέτρηση, είναι να διαθέτουμε δύο τύπους του ίδιου ψυχομετρικού εργαλείου, το οποίο και χορηγούμε ως ενιαίο σύνολο σε όλους τους συμμετέχοντες. Οι δύο αυτοί τύποι θα πρέπει, ασφαλώς, να είναι ισοδύναμοι, δηλαδή να εμπεριέχουν τις ίδιες ψυχολογικές έννοιες, να αναφέρονται στις ίδιες μεταβλητές, να αποτελούν, ουσιαστικά, παραλλαγή ο ένας του άλλου. Έχοντας αυτούς τους ισοδύναμους τύπους έτοιμους, μπορούμε να τους διατάξουμε σε ευθεία σειρά για τους μισούς συμμετέχοντες και σε αντίστροφη σειρά για τους άλλους μισούς (για να αποφύγουμε επιδράσεις σειράς), να χορηγήσουμε τις δύο αυτές μορφές και να υπολογίσουμε τον δείκτη συνάφειας για τις μετρήσεις που προκύπτουν. Οι μετρήσεις αυτές, δεν αναφέρονται σε ένα προς ένα τα ερωτήματα των δύο εναλλακτικών μορφών του τεστ, αλλά στις τελικές τιμές που προκύπτουν. Οι τιμές αυτές είναι συνήθως τα αθροίσματα των επιμέρους ερωτημάτων, αλλά αυτό εξαρτάται και από τη φύση και την τεχνική που εμπεριέχει το τεστ (μπορεί για κάποιες ερωτήσεις το βάρος-συντελεστής να είναι διαφορετικός)· ο δείκτης συνάφειας όμως, υπολογίζεται για τις τελικές αυτές τιμές, διότι αυτές εκφράζουν *συνολικά* τους δύο ισοδύναμους τύπους.

Ο έλεγχος μέσω της αξιοπιστίας ισοδύναμων τύπων επιχειρεί να θεραπεύσει το θόρυβο που προκαλείται εξαιτίας του περιεχομένου του τεστ, δηλαδή της συγκεκριμένης θεματολογίας στην οποία αυτό αναφέρεται. Τίθεται, δηλαδή, το ερώτημα του κατά πόσο τα αποτελέσματά μας δεν θα είναι σταθερά, επειδή υπάρχουν και χιλιάδες άλλοι τρόποι αξιολόγησης των ίδιων εννοιών. Εάν καταφέρουμε να δείξουμε ότι δύο ισοδύναμοι τρόποι αξιολόγησης μιας ψυχολογικής έννοιας καταλήγουν στα ίδια αποτελέσματα, τότε μπορούμε να ισχυριστούμε ότι ο θόρυβος αυτού του τύπου στο τεστ αυτό είναι περιορισμένος.

Ασφαλώς, υπάρχουν πρακτικά προβλήματα. Ποιος από τους δυο ισοδύναμους τύπους θα πρέπει να αποτελέσει την τελικά επιλεγόμενη ψυχομετρική κλίμακα αξιολόγησης; Πώς μπορούμε να αποδείξουμε ότι οι δύο ισοδύναμοι τύποι είναι πράγματι «ισοδύναμοι»; Πόσο εύκολο είναι να δημιουργήσουμε όχι ένα, αλλά δύο σύνολα ερωτημάτων-δοκιμασιών που αναφέρονται στη μέτρηση της ίδιας έννοιας; Τα προβλήματα αυτά, μαζί με άλλες τεχνικές δυσκολίες, οδηγούν τους ερευνητές στην ψυχομετρία να εφαρμόζουν συνήθως άλλες μεθόδους ελέγχου της αξιοπιστίας μιας ψυχομετρικής κλίμακας αξιολόγησης.

Αξιοπιστία των ημικλάστων

Καθώς η δημιουργία ενός ψυχομετρικού εργαλείου είναι ιδιαίτερα επίπονη και χρονοβόρα, και καθώς το να δημιουργήσει κανείς έναν τύπο ερωτηματολογίου ή οποιασδήποτε μορφής και δομής ψυχομετρική διαδικασία είναι πράξη που απαιτεί γνώσεις, ευελιξία, μελέτη, υπομονή, αυστηρότητα, διαδοχικές έρευνες-πλότους, συνεχή επαναπροσδιορισμό των εννοιών και των λειτουργικών ορισμών τους, κ.ο.κ. (ο κατάλογος αυτός περιέχει αναρίθμητες «ικανότητες» που πρέπει να έχει ο εμπνευστής και δημιουργός ενός τεστ), είναι, σαφέστατα, χρονικά και ερευνητικά προτιμότερο κάποιος να μην επιχειρήσει να δημιουργήσει δεύτερο, ισοδύναμο τύπο του τεστ. Και πάλι όμως, θα πρέπει να είμαστε σε θέση να εκτιμήσουμε την αξιοπιστία της ψυχομετρικής κλίμακας που έχουμε δημιουργήσει και να ελέγξουμε τον παράγοντα θορύβου που προκαλείται από το ίδιο το περιεχόμενο του συγκεκριμένου τεστ. Αντί, όμως, να έχουμε δύο συγκρινόμενους τύπους του ίδιου τεστ, διαιρούμε το τεστ και συγκρίνουμε μεταξύ τους τα δύο τμήματά του.

Έχουμε ήδη αναφέρει ότι, προοδευτικά, σε μία ψυχομετρική κλίμακα αξιολόγησης, οι ερωτήσεις προχωρούν από τις ευκολότερες στις δυσκολότερες (ειδικά στα τεστ ικανοτήτων και νοημοσύνης), κατά συνέπεια, θα ήταν τελείως εσφαλμένο να διαιρέσουμε το τεστ τέμνοντάς το στη μέση, διότι έτσι θα είχαμε στο πρώτο μέρος μόνο τις εύκολες και στο δεύτερο μέρος μόνο τις δύσκολες ερωτήσεις. Για να διαιρέσουμε το τεστ θα επιλέξουμε τις ερωτήσεις με περιττό αύξοντα αριθμό και θα τις τοποθετήσουμε στη μία ομάδα ερωτημάτων και τις ερωτήσεις με άρτιο αύξοντα αριθμό θα τις τοποθετήσουμε στην άλλη ομάδα ερωτημάτων¹⁰, ισοκατανέμοντας έτσι τον δείκτη δυσκολίας των ερωτημάτων. Αυτές οι δύο ομάδες ερωτημάτων καταλήγουν, αθροίζοντας τα επιμέρους ερωτήματα, σε δύο τιμές της ίδιας έννοιας και αυτές οι τιμές είναι διαθέσιμες για κάθε συμμετέχοντα του δείγματός μας. Θεωρητικά, τα δύο αυτά τμήματα θα πρέπει να παρουσιάζουν συνάφεια ως προς τις τιμές τους. Μπορούμε, επομένως, σε πρώτο στάδιο, να υπολογίσουμε τον δείκτη αυτόν συνάφειας, ο οποίος στην ουσία είναι ο δείκτης αξιοπιστίας των ημικλάστων του τεστ.

Η αξιοπιστία μιας οποιασδήποτε ψυχομετρικής κλίμακας εξαρτάται, σε πολύ μεγάλο βαθμό, από τον αριθμό των ερωτημάτων που χρησιμοποιούνται στην κλίμακα αυτή. Διαιρώντας το τεστ παραβιάζουμε τον κανόνα αυτόν, διότι κατ' ουσίαν πλέον διαθέτουμε δύο τεστ της μισής από την πραγματική έκτασης. Χρειάζεται επομένως να διορθώσουμε ως προς την αρχική έκταση της κλίμακας τον δείκτη συνάφειας στον οποίο έχουμε καταλήξει και ο οποίος αρχικά εξέφραζε την αξιοπιστία των ημικλάστων. Η διόρθωση αυτή γίνεται μέσω του υποδείγματος των *Spearman & Brown*:

$$r_{sh} = \frac{2 \times r_h}{1 + r_h} \quad \left[\begin{array}{l} \text{όπου: } r_h \text{ είναι ο αρχικός δείκτης συνάφειας μεταξύ των ημικλάστων, και} \\ r_{sh} \text{ είναι ο τελικός δείκτης αξιοπιστίας των ημικλάστων του τεστ,} \\ \text{διορθωμένος ως προς την αρχική έκταση του τεστ.} \end{array} \right]$$

¹⁰ Υπάρχουν πολλοί τρόποι να διαιρέσει κανείς ένα σύνολο ερωτημάτων-δοκιμασιών. Σύμφωνα με τη *Συνδυαστική θεωρία*, οι συνδυασμοί v αντικειμένων σε ομάδες τις τάξεως μ , ορίζονται ως ακολούθως:

$$\binom{v}{\mu} = \frac{v!}{(v-\mu)! \mu!} \quad \text{όπου, } v! = v(v-1)(v-2)(v-3)\dots(v-v+1). \text{ Κατά συνέπεια, εάν σε μία ψυχομετρική}$$

κλίμακα αξιολόγησης υπάρχουν 20 δοκιμασίες και θέλουμε να ορίσουμε δύο ομάδες της τάξεως των 10 δοκιμασιών, τότε οι διάφοροι συνδυασμοί τέτοιων 10άδων ανέρχονται στους 184.756.

Η πρώτη πληροφορία που λαμβάνουμε από τον δείκτη αυτόν είναι το κατά πόσο το περιεχόμενο της κλίμακας που χρησιμοποιούμε συμβάλλει στη σταθερότητα των τιμών που προκύπτουν από την μέτρηση ενός ή περισσότερων ατόμων. Εάν, για παράδειγμα, ο δείκτης της αξιοπιστίας των ημικλάστων, διορθωμένος κατά *Spearman & Brown*, είναι 0,76, αυτό σημαίνει ότι 76 φορές στις 100 οι μετρήσεις μας θα οφείλονται στο περιεχόμενο του τεστ και μόνο. Επίσης, ο ίδιος δείκτης σημαίνει ότι 76 στις 100 φορές που θα χορηγούσαμε την ίδια ψυχομετρική κλίμακα αξιολόγησης στο ίδιο άτομο θα λαμβάναμε τις ίδιες απαντήσεις. Εκτός όμως των πληροφοριών αυτών, μπορούμε, χρησιμοποιώντας τον δείκτη αξιοπιστίας των ημικλάστων, να καταλήξουμε και σε περαιτέρω συμπεράσματα για την κλίμακα και τις μετρήσεις που προκύπτουν από αυτήν, συμπεράσματα ιδιαίτερα σημαντικά για τη χρησιμότητα του τεστ και την γενικότερη δυνατότητα ερμηνείας των μετρήσεων που προκύπτουν από αυτό.

Τίθεται το ερώτημα: Ας υποθέσουμε ότι, μετρώντας τις μηχανικές ικανότητες ενός ατόμου μέσω ενός ψυχομετρικού εργαλείου που έχουμε κατασκευάσει, καταλήγουμε σε μια μέτρηση, έστω 30. Μετρώντας, με το ίδιο εργαλείο τις μηχανικές ικανότητες ενός άλλου ατόμου, καταλήγουμε στην τιμή 36. Διαφέρουν μεταξύ τους τα δύο αυτά άτομα, ως προς τις μηχανικές ικανότητες; Ποια είναι, δηλαδή, τα όρια που διαχωρίζουν, με βάση πάντα το περιεχόμενο της ψυχομετρικής κλίμακας, τις διάφορες βαθμίδες ικανότητας και πώς μπορούμε να εκφράσουμε τα όρια αυτά αριθμητικά, ώστε να μπορούμε να συγκρίνουμε τις τελικές τιμές που προκύπτουν; Η απάντηση δίνεται από τον δείκτη αξιοπιστίας των ημικλάστων, με βάση τον οποίο είμαστε σε θέση να υπολογίσουμε, για το συγκεκριμένο τεστ, τα διαστήματα εμπιστοσύνης εκατέρωθεν της οποιασδήποτε τιμής προκύψει από μια εξέταση. Μπορούμε, δηλαδή, να εκτιμήσουμε τα όρια μέσα στα οποία, λόγω της παραμέτρου του δειγματοληπτικού σφάλματος, θα κυμαίνεται 95 (ή 99) φορές στις 100, ή 999 φορές στις 1000, η τιμή που προκύπτει από τη συγκεκριμένη αξιολόγηση. Με τον τρόπο αυτόν, μπορούμε να εκτιμήσουμε μέσα σε ποια όρια θα μπορούσε να κυμαίνεται η τιμή που έχει προκύψει για τον εξεταζόμενο 'X' και γνωρίζοντας τα όρια αυτά θα γνωρίζουμε ότι, όποια άλλη μέτρηση άλλου ατόμου βρίσκεται εντός των ορίων, στην ουσία δεν διαφέρει από την τιμή που αντιστοιχεί στον εξεταζόμενο 'X'.

Από τη θεωρία των αληθών τιμών προκύπτει το ακόλουθος υπόδειγμα:

$$\text{Διασπορά σφάλματος} = \text{διασπορά} \otimes (1 - r_{sh}), \text{ δηλαδή } s^2 = s^2 \otimes (1 - r_{sh})$$

Από το υπόδειγμα αυτό, εξάγοντας την τετραγωνική ρίζα κάθε όρου για να εξαλείψουμε τα τετράγωνα, καταλήγουμε στο ακόλουθο υπόδειγμα:

$$\begin{aligned} \text{Τυπικό σφάλμα μέτρησης} &= \text{τυπική απόκλιση} \otimes \sqrt{(1 - r_{sh})}, \text{ δηλαδή} \\ s_e &= s \otimes \sqrt{(1 - r_{sh})} \end{aligned}$$

Γνωρίζοντας, επομένως, τον δείκτη αξιοπιστίας των ημικλάστων, διορθωμένο για την έκταση της κλίμακας, όπως αυτός προέκυψε από τη δειγματοληπτική διαδικασία, καθώς και την τυπική απόκλιση της συγκεκριμένης μέτρησης, στην οποία καταλήγει το τεστ και, η οποία μας είναι διαθέσιμη μέσω, και πάλι, της δειγματοληπτικής διαδικασίας, μπορούμε να υπολογίσουμε το τυπικό σφάλμα για τη συγκεκριμένη αυτή μέτρηση. Έχοντας υπολογίσει το σφάλμα αυτό, είμαστε πλέον σε θέση να υπολογίσουμε το διάστημα εμπιστοσύνης για οποιαδήποτε τιμή προκύπτει από το τεστ χρησιμοποιώντας τα θεωρήματα των δειγματοληπτικών κατανομών και συγκεκριμένα:

$$\text{Όρια εμπιστοσύνης} = \text{αρχική τιμή μέτρησης} \pm (1,96 \otimes s_e)$$

Χρησιμοποιούμε, στην περίπτωση, τον αριθμό 1,96 διότι αυτή είναι η κρίσιμη δικατάληκτη z τιμή που σύμφωνα με τη θεωρία των δειγματοληπτικών κατανομών και τη θεωρητική z κατανομή, αποκόπτει το ακραίο 5% των τιμών της κατανομής αφήνοντας μέσα στα όρια εμπιστοσύνης το 95%. Εάν θέλουμε να αναφερθούμε σε όρια εμπιστοσύνης που αφορούν στο 99%, η αντίστοιχη κρίσιμη z τιμή που θα χρησιμοποιήσουμε είναι η 2,58. Ο λόγος για τον οποίο χρησιμοποιούνται οι κρίσιμες z-τιμές αφορά και πάλι σε διόρθωση στην οποία προβαίνουμε σχετικά με την κλίμακα μέτρησης την οποία έχουμε επιλέξει να χρησιμοποιήσουμε. Έτσι, μια δίτιμη κλίμακα μέτρησης (0 και 1) για 30 ερωτήματα καταλήγει σε minimum 0 και maximum 30. Για να εκφράσουμε την κλίμακα αυτή μέτρησης σε σχέση με τη θεωρητική z-κατανομή, έκφραση η οποία επιβάλλεται για

τον καθορισμό των ορίων εμπιστοσύνης, και ταυτόχρονα να καταλήξουμε σε αριθμούς συμβατούς με την αρχική κλίμακα μέτρησης, πολλαπλασιάζουμε με την κρίσιμη z-τιμή η οποία ορίζει το ποσοστό βεβαιότητας το οποίο έχουμε εκ των προτέρων καθορίσει (95%, 99%, κ.ο.κ.).

Ας δούμε ένα ολοκληρωμένο παράδειγμα. Υποθέτουμε ότι έχουμε δημιουργήσει μια ψυχομετρική κλίμακα μέτρησης του έκδηλου άγχους και τη χορηγήσαμε δειγματοληπτικά σε 200 συμμετέχοντες. Η κλίμακα αυτή περιλαμβάνει 38 επιμέρους ερωτήματα τα οποία ορίζονται εννοιολογικά και λειτουργικά ως επιμέρους μετρήσεις διαφόρων πτυχών του έκδηλου άγχους και βαθμολογούνται από τους συμμετέχοντες επί τη βάση μιας 5βαθμης κλίμακας *Likert*. Έχοντας συγκεντρώσει τις μετρήσεις από τα 200 άτομα, διαιρούμε την κλίμακα σε δύο τμήματα (στην περίπτωση, ίσα) των 19 ερωτήσεων, επιλέγοντας για το ένα τμήμα, με βάση τον αύξοντα αριθμό τους, τις άρτιες ερωτήσεις και για το άλλο τις περιττές. Αθροίζουμε τις επιμέρους μετρήσεις και καταλήγουμε με δύο μετρήσεις έκδηλου άγχους, μία για κάθε τμήμα της κλίμακας, για κάθε συμμετέχοντα. Υπολογίζουμε τον δείκτη *Pearson r* για τις τιμές αυτές, ο οποίος προκύπτει, έστω, ίσος με 0,57. Διορθώνοντας τον δείκτη αυτόν αξιοπιστίας ημικλάστων για την αρχική έκταση της κλίμακας (38 ερωτήματα) προκύπτει ο τελικό δείκτης r_{sh} ο οποίος είναι περίπου ίσος με 0,73. Από τον δείκτη αυτόν και την τυπική απόκλιση που προκύπτει από την κατανομή συχνότητας των τελικών μετρήσεων (δηλαδή από την κατανομή συχνότητας για το συνολικό άθροισμα των 38 ερωτημάτων, για κάθε συμμετέχοντα), και η οποία στην περίπτωση είναι, έστω, περίπου 31, μπορούμε να υπολογίσουμε το τυπικό σφάλμα της μέτρησης: $31 \otimes \sqrt{1-0,73} \approx 16,11$.

Στη συνέχεια, μπορούμε να προχωρήσουμε σε υπολογισμό διαστημάτων εμπιστοσύνης για συγκεκριμένες πλέον τελικές τιμές μέτρησης που μπορεί να προκύψουν για κάποιον συμμετέχοντα. Οι τιμές που μπορεί να προκύψουν, με βάση την κλίμακα βαθμολόγησης των ερωτημάτων, η οποία έχει *maximum* βαθμίδα το '5' και *minimum* βαθμίδα το '1', κυμαίνονται από *minimum* $38 \otimes 1 = 38$ μέχρι *maximum* $38 \otimes 5 = 190$. Θα μπορούσε, λοιπόν, θεωρητικά, σε κάποια μελλοντική μέτρηση, να προκύψει μια τελική τιμή έκδηλου άγχους για κάποιον εξεταζόμενο (Α) ίση με 125 και, για κάποιον άλλο εξεταζόμενο (Β), τελική τιμή έκδηλου άγχους ίση με 144. Το

ερώτημα είναι, αν αυτές οι δύο τελικές τιμές διαφέρουν μεταξύ τους, δηλαδή αν ο εξεταζόμενος (B) με την τιμή 144 έχει περισσότερο άγχος από αυτόν με την τιμή 125. Υπολογίζοντας τα όρια εμπιστοσύνης για όποια από τις δύο τιμές, έστω την 125, βρίσκουμε ότι (95 φορές στις 100) ισχύει: $125 \pm (1,96 \otimes 16,11) = 125 \pm 31,58$. Δηλαδή, τα όρια αυτά εκτείνονται 95 φορές στις 100 μεταξύ 93,42 και 156,58. Με άλλα λόγια, κάποιος εξεταζόμενος (X) που έχει βαθμό έκδηλου άγχους 90 διαφέρει από αυτόν (A) που έχει 125 ενώ ο (A) διαφέρει από κάποιον (Ψ) που έχει βαθμό έκδηλου άγχους τουλάχιστον 157. Έτσι, οι δύο εν λόγω εξεταζόμενοι (A=125 & B=144) δεν διαφέρουν ως προς το έκδηλο άγχος τους, δηλαδή παρουσιάζουν τον ίδιο βαθμό έκδηλου άγχους αν και διαφέρουν αριθμητικά (144 – 125) 19 μονάδες έκδηλου άγχους. Κάποιος, επίσης, με βαθμό έκδηλου άγχους 153 δεν διαφέρει από εκείνον που έχει βαθμό έκδηλου άγχους 125 ούτε και από κάποιον με βαθμό 175. Διαφέρει όμως από κάποιον με βαθμό έκδηλου άγχους 121 ή λιγότερο.

Όπως γίνεται κατανοητό από το παράδειγμα, όταν ένα τεστ δεν μπορεί να διαφοροποιήσει μεταξύ τιμών που απέχουν αριθμητικά πολύ (σε σχέση πάντα με το θεωρητικό εύρος τιμών), επειδή η διασπορά των τιμών είναι μεγάλη και επειδή η αξιοπιστία των ημικλάστων είναι σχετικά μικρή, το τεστ αυτό θεωρείται αναξιόπιστο, ενώ οι πληροφορίες που μας παρέχει είναι ασαφείς και αδιαφοροποίητες μεταξύ τους. Φανταστείτε, για ένα τεστ νοημοσύνης, να προκύπτουν δύο τελικές τιμές (προσοχή, εδώ δεν αναφερόμαστε σε δείκτες IQ, αλλά απλά, σε τελικά αθροίσματα επιμέρους ερωτημάτων), οι οποίες απέχουν μεταξύ τους 30 μονάδες, ενώ το συνολικό θεωρητικό εύρος τιμών είναι 150, και να μην μπορούμε λόγω του δείκτη αξιοπιστίας ημικλάστων και της διασποράς, να διαφοροποιήσουμε μεταξύ των τιμών αυτών, δηλαδή να είμαστε υποχρεωμένοι να δεχτούμε ότι οι δύο αυτές τιμές, αν και απέχουν αριθμητικά πολύ, εκφράζουν ίδια νοημοσύνη.

Ο δείκτης αξιοπιστίας των ημικλάστων είναι πολύ χρήσιμος και χρησιμοποιείται ευρύτατα για να εξαχθούν πληροφορίες σχετικά με το κατά πόσο το περιεχόμενο μιας κλίμακας αποδίδει αληθείς, με μικρά περιθώρια σφάλματος, τιμές. Υπάρχουν, όμως, κάποιες δυσκολίες. Όταν μια ψυχομετρική κλίμακα απαρτίζεται από μικρό αριθμό ερωτημάτων, η διόρθωση των *Spearman & Brown* τίθεται υπό αμφισβήτηση, καθώς, εάν η πραγματική έκταση της κλίμακας είναι 6 ερωτήματα και υπολογίσουμε τον δείκτη συνάφειας μεταξύ των δύο ημικλάστων τριάδων, ο δείκτης

αυτός δεν μπορεί να «διορθωθεί», στην ουσία με πλασματικό τρόπο, σαν να ήταν η πραγματική έκταση της κλίμακας 30 ερωτήματα. Μια εναλλακτική λύση στον δείκτη αξιοπιστίας που προκύπτει από την διόρθωση *Spearman-Brown*, είναι ο δείκτης αξιοπιστίας ημικλάστων του *Guttman*, ο οποίος δεν υποθέτει εξ αρχής ότι τα δύο μισά έχουν την ίδια αξιοπιστία ή την ίδια διασπορά τιμών και παρακάμπτει έτσι το θέμα της διόρθωσης για την αρχική έκταση της κλίμακας. Το πρόβλημα, όμως, εδώ είναι ότι καταλήγουμε σε δύο δείκτες αξιοπιστίας ημικλάστων για την ψυχομετρική κλίμακα, έναν για κάθε μισό της, γεγονός που καθιστά τους δείκτες αυτούς δύσχρηστους.

Ένα άλλο σημείο, που χρήζει ιδιαίτερης μνείας, αφορά στο γεγονός ότι η εκτιμητική διαδικασία διαστημάτων εμπιστοσύνης μέσω του δείκτη αξιοπιστίας ημικλάστων αναφέρεται στις πρωτογενείς τιμές και όχι σε δευτερογενείς μετατροπές τους που γίνονται με βάση τις θεωρητικές δειγματοληπτικές κατανομές. Συνήθως όμως, στα ψυχομετρικά τεστ συγκρίνουμε δύο «επιδόσεις» με βάση τις δευτερογενείς τιμές οι οποίες τους αντιστοιχούν. Επίσης, οι πιθανοί δείκτες αξιοπιστίας ημικλάστων για ένα σύνολο δοκιμασιών-ερωτημάτων είναι πάρα πολλοί, και εξαρτώνται εν μέρει από το πώς θα διαιρεθεί η αρχική ψυχομετρική κλίμακα αξιολόγησης στα δύο ημίση. Κατά συνέπεια, υπό συνθήκες, ο δείκτης αξιοπιστίας ημικλάστων, ίσως δεν πρέπει να χρησιμοποιηθεί, αλλά στη θέση του να αναζητηθεί εναλλακτικός δείκτης αξιοπιστίας της ψυχομετρικής κλίμακας, όπως ο δείκτης αξιοπιστίας εσωτερικής συνέπειας, ο οποίος μάλιστα ελέγχει τις παραμέτρους λάθους που προξενούνται από το περιεχόμενο της κλίμακας αλλά και από την ετερογένεια των επιμέρους ερωτημάτων της. Ο δείκτης αυτός, που αποτελεί τον μέσο όρο όλων των πιθανών δεικτών αξιοπιστίας ημικλάστων για μια ψυχομετρική κλίμακα αξιολόγησης, καθώς και η διαδικασία υπολογισμού και ερμηνείας του αναλύονται αμέσως παρακάτω.

Αξιοπιστία εσωτερικής συνέπειας

Ο υπολογισμός, η ερμηνεία και η γενικότερη χρήση του δείκτη αξιοπιστίας εσωτερικής συνέπειας στηρίζεται στη θεωρία των δειγματοληπτικών κατανομών. Η έννοια της αξιοπιστίας αυτής, σε σχέση με τα υπόλοιπα είδη αξιοπιστίας, προέρχεται από ένα διαφορετικό θεωρητικό υπόβαθρο, τη «θεωρία της δυνατότητας γενίκευσης»

(*generalizability theory*). Κυριότεροι εκπρόσωποι της διαφορετικής αυτής θεώρησης είναι οι *Kuder & Richardson*, και ο *L. J. Cronbach*. Ο τελευταίος, μαζί με τους συνεργάτες του, υποστήριξε ότι η διάκριση μεταξύ αξιοπιστίας και εγκυρότητας ενός τεστ είναι εσφαλμένη και ότι αυτό που ενδιαφέρει είναι η δυνατότητα γενίκευσης μέσω του τεστ. Το βασικό λοιπόν ερώτημα για την αξιοπιστία εσωτερικής συνέπειας είναι το κατά πόσο οι μετρήσεις ενός τεστ που έχουν προκύψει από ένα δείγμα συμμετεχόντων, θα εξακολουθούν να παρατηρούνται σε μεγάλο βαθμό ομοιότητας σε άπειρα παρόμοια δείγματα. Για τη θεώρηση της δυνατότητας γενίκευσης, είναι πολύ σημαντικό οι μετρήσεις που προκύπτουν για ένα δείγμα να μπορούν να γενικευθούν στον αντίστοιχο στατιστικό πληθυσμό, ενώ κάθε ερώτημα-δοκιμασία θεωρείται αυτόνομη ψυχομετρική αξιολόγηση η οποία ανήκει σε ένα γενικότερο σύνολο-δειγματοχώρο από πιθανά ερωτήματα-δοκιμασίες. Για τους λόγους αυτούς, η αξιοπιστία εσωτερικής συνέπειας είναι άμεσα συνυφασμένη με τις θεωρητικές κατανομές, όπως περιγράφονται από τη θεωρία των πιθανοτήτων και από την εκτιμητική θεωρία της στατιστικής, καθώς και με τα θεωρήματα που προκύπτουν από τις δειγματοληπτικές κατανομές.

Οι δειγματοληπτικές κατανομές, ως γνωστόν, αναφέρονται σε άπειρα, τυχαία, ισοπληθή δείγματα, από τα οποία προκύπτουν μετρήσεις για κάποια μεταβλητή ενδιαφέροντος. Στην ίδια βάση, ο *Cronbach* θέλησε να εκτιμήσει, μέσω των μετρήσεων που προκύπτουν από ένα ψυχομετρικό εργαλείο, την πιθανότητα οι μετρήσεις αυτές να επαναπροκύπτουν από άπειρα, τυχαία, ισοπληθή με το αρχικό, δείγματα και να εκφράσει την πιθανότητα αυτή, που έχει ποσοστιαία μορφή, ως δείκτη αξιοπιστίας. Η πιθανότητα αυτή είναι πράγματι δείκτης αξιοπιστίας, καθώς μας πληροφορεί για το πόσο σταθερά θα είναι τα αποτελέσματα ενός ψυχομετρικού τεστ, εάν επαναλάβουμε τις μετρήσεις άπειρες φορές· εφόσον στηρίζεται στις ιδιότητες των τυχαίων, ισοπληθών δειγμάτων, παραπέμπει άμεσα στις θεωρητικές κατανομές, επομένως στο στατιστικό πληθυσμό, επομένως και στην αξιοπιστία που αφορά στην ατομική ψυχομετρική διαδικασία. Άλλωστε, η ψυχολογία έχει τελικό της στόχο την προσέγγιση των ατομικών χαρακτηριστικών, έχει δηλαδή επίκεντρό της το άτομο, απλά η επικέντρωσή της στον στόχο αυτόν γίνεται μέσω των δειγματοληπτικών μεθόδων.

Ο υπολογισμός του δείκτη αξιοπιστίας εσωτερικής συνέπειας προκύπτει είτε από τον τύπο KR-20 (*Kuder-Richardson*), όταν τα ερωτήματα της ψυχομετρικής κλίμακας αξιολόγησης επιδέχονται μόνο δύο απαντήσεις (σωστή ή λάθος απάντηση), όταν δηλαδή τα στοιχεία που αναλύονται είναι διωνυμικού τύπου (δίτιμες μετρήσεις) είτε μέσω του τύπου *Cronbach α*, όπου το α είναι ο δείκτης αξιοπιστίας. Ο υπολογιστικός τύπος είναι:

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum s^2}{s_T^2} \right)$$

όπου: k είναι ο αριθμός των ερωτημάτων,
 $\sum s^2$, το άθροισμα των τετραγωνισμένων τυπικών αποκλίσεων κάθε ερωτήματος, και
 s_T^2 , η διασπορά της κατανομής συχνότητας για τις τελικές μετρήσεις του τεστ (αθροίσματα).

Ο δείκτης *Cronbach α* χρησιμοποιείται για την ανεύρεση της αξιοπιστίας μιας ψυχομετρικής κλίμακας αξιολόγησης, όταν τα επιμέρους ερωτήματα της κλίμακας αυτής επιδέχονται βαθμολόγησης με περισσότερες από δύο βαθμίδες διαφοροποίησης (π.χ. 5βαθμη *Likert* κλίμακα μέτρησης). Μια δεύτερη, πολύ σημαντική, προϋπόθεση είναι ότι τα ερωτήματα της κλίμακας πρέπει να είναι τουλάχιστον τρία σε αριθμό. Με δύο ερωτήματα στην κλίμακα αξιολόγησης, δεν μπορεί να υπολογισθεί ο δείκτης *Cronbach α*, ενώ μπορούμε να υπολογίσουμε μόνο τον δείκτη συνάφειας μεταξύ των δύο αυτών ερωτημάτων.

Ενώ το υπολογιστικό υπόδειγμα αναφέρεται σε διακυμάνσεις τιμών και σε αθροίσματα διακυμάνσεων, κατ' ουσίαν βασίζεται στον πίνακα συναφειών μεταξύ όλων των πιθανών ζευγών ερωτημάτων του τεστ. Εφόσον παρατηρείται συνοχή και ομοιομορφία, δηλαδή *ομοιογένεια* των δεικτών συνάφειας μέσα στον πίνακα αυτόν, και εφόσον οι δείκτες αυτοί είναι αρκετά υψηλοί, ο δείκτης *Cronbach α* αναμένεται να είναι ικανοποιητικός. Ο πίνακας συναφειών δεν θα πρέπει να έχει πολύ υψηλές συνάφειες σε όλο του το εύρος. Εάν συμβαίνει αυτό, η ερμηνεία του πίνακα είναι ότι τα ερωτήματα αλληλεπικαλύπτονται και δεν χρειάζονται όλα, επειδή πολλά από αυτά μετρούν τα ίδια ακριβώς πράγματα. Στην περίπτωση μάλιστα αυτή, ο δείκτης αξιοπιστίας εσωτερικής συνέπειας θα καταλήξει να είναι είτε υπερβολικά υψηλός (π.χ. $\alpha = 1,00$) είτε εκτός των ορίων. Τα όρια του δείκτη ξεκινούν από το $-1,00$ και φτάνουν στο $+1,00$, ενώ ερμηνεύονται με τον ίδιο τρόπο που ερμηνεύεται ο δείκτης *Pearson r*, αν και στην περίπτωση δεν μας ενδιαφέρει το πρόσημο του α . Έτσι, αν ο

δείκτης α προκύψει να είναι $-0,78$, αυτό σημαίνει ότι 78 φορές στις 100 που θα επαναχορηγήσουμε το ίδιο ψυχομετρικό εργαλείο στο ίδιο άτομο, θα λάβουμε παρόμοιες μετρήσεις. Εάν ο δείκτης, όπως αναφέρθηκε πιο πριν, είναι εκτός ορίων, έστω $3.524,69$, αυτό θα σημαίνει ότι, κατά την υπολογιστική διαδικασία, παραβιάστηκαν οι στατιστικές παραδοχές που αφορούν στον πίνακα συναφειών των ερωτημάτων της κλίμακας και ότι υπάρχει πρόβλημα ως προς τις αλληλεξαρτήσεις των ερωτημάτων που χρησιμοποιούμε. Αν τέλος, προκύψει ένας δείκτης α μικρότερος από την απόλυτη τιμή $0,70$, ίσως θα πρέπει να επανεξετάσουμε το θέμα της εσωτερικής συνέπειας των ερωτημάτων της κλίμακάς μας.

Ο δείκτης αξιοπιστίας εσωτερικής συνέπειας είναι πολύ ευαίσθητος στις παραμέτρους μεγέθους του δείγματος από το οποίο προκύπτει καθώς και στον αριθμό των ερωτημάτων που περιλαμβάνει η ψυχομετρική κλίμακα. Όσα περισσότερα άτομα έχουμε στο δείγμα από το οποίο έχουμε συλλέξει τις μετρήσεις μας, τόσο πιο μεγάλη σταθερότητα παρουσιάζει ο δείκτης αξιοπιστίας εσωτερικής συνέπειας. Ακόμη και σ' αυτή την περίπτωση όμως, εάν η ψυχομετρική κλίμακα αξιολόγησης περιλαμβάνει πολύ λίγα ερωτήματα (έστω 5), είναι πολύ δύσκολο ο δείκτης α να είναι ικανοποιητικός ως προς το βαθμό.

Η επίδραση του αριθμού των συμμετεχόντων στην υπολογιστική διαδικασία του δείκτη *Cronbach α* γίνεται προφανής εάν αναλογιστούμε την επίδραση του μεγάλου μεγέθους του δείγματος στο τυπικό σφάλμα μέτρησης των στατιστικών δεικτών για κάθε ερώτημα-δοκιμασία. Όσο μεγαλύτερο το N , τόσο μικρότερο το τυπικό αυτό σφάλμα για κάθε ερώτημα, επομένως τόσο μικρότερη και η διασπορά του. Αθροίζοντας αυτές τις διασπορές, ο αριθμητής του κλάσματος στο υπολογιστικό υπόδειγμα αξιοπιστίας εσωτερικής συνέπειας ελαχιστοποιείται, ενώ δεν συμβαίνει το ίδιο για τη διασπορά του αθροίσματος όλων των επιμέρους ερωτημάτων. Έτσι, το τελικό κλάσμα ελαχιστοποιείται, μεγιστοποιώντας, πιθανώς πλασματικά, τον τελικό δείκτη. Αντίστροφα, ένα πολύ μικρός αριθμός συμμετεχόντων μεγιστοποιεί το τυπικό σφάλμα μέτρησης για κάθε ερώτημα, μεγιστοποιώντας έτσι τον αριθμητή του κλάσματος στην υπολογιστική διαδικασία και ελαχιστοποιώντας τον τελικό δείκτη. Ο αριθμός των ερωτημάτων είναι επίσης σημαντική παράμετρος, καθώς ένα μεγάλο σύνολο ερωτημάτων περιέχει σε κάθε ερώτημα κάποιον μικρό αριθμό πληροφοριών που στο τελικό άθροισμα των τιμών καταλήγει σε συνολικά μεγάλη διασπορά τιμών,

ενώ ακόμη και μεγάλα τυπικά σφάλματα μέτρησης για κάθε ερώτημα δεν επαρκούν για να ισορροπήσουν αναλογικά τον αριθμητή του κλάσματος.

Ασφαλώς, δεν είναι δυνατό να δημιουργούμε και να χορηγούμε ψυχομετρικές κλίμακες αξιολόγησης που περιλαμβάνουν 400 ή 500 ερωτήματα, διότι υπάρχουν και οι επιδράσεις κόπωσης που υφίστανται οι εξεταζόμενοι, επομένως, θα πρέπει να περιλαμβάνουμε στην κλίμακα τόσα ερωτήματα ώστε να μην καταστρατηγείται η λογική του *Cronbach* και, ταυτόχρονα, να μην καταλήγουμε σε τερατώδεις κλίμακες αξιολόγησης που χρειάζονται ώρες για να απαντηθούν. Ένας γενικότερος κανόνας, αν και δεν είναι απόλυτος, είναι ότι θα χρησιμοποιήσουμε τόσα ερωτήματα και τόσους συμμετέχοντες, όσα και όσους θα χρησιμοποιούσαμε, αν το αντικείμενό μας ήταν μια ερευνητική διαδικασία μέσω του ψυχομετρικού αυτού εργαλείου και όχι η προσπάθεια ανεύρεσης της αξιοπιστίας του. Ο στόχος μας δηλαδή, θα πρέπει να είναι να πληρούμε, όσο το δυνατόν, τις μεθοδολογικές και στατιστικές απαιτήσεις και παραδοχές.

Μεταξύ βαθμολογητών αξιοπιστία

Σε πολλές περιπτώσεις, όταν τίθεται ζήτημα υποκειμενικής βαθμολόγησης των απαντήσεων των συμμετεχόντων σε ένα τεστ, είναι ανάγκη να υπολογίσουμε την αξιοπιστία μεταξύ των βαθμολογητών. Όλοι οι προηγούμενοι δείκτες αξιοπιστίας αναφέρονται σε περιπτώσεις όπου οι απαντήσεις του εξεταζομένου βαθμολογούνται με βάση αντικειμενικά, ή έστω κοινά αποδεκτά και συμφωνημένα, κριτήρια. Όταν όμως θέλουμε, π.χ., να βαθμολογήσουμε ως προς την ικανότητα λόγου τις εκθέσεις ιδεών των συμμετεχόντων, τότε υπάρχει πρόβλημα αντικειμενικότητας της βαθμολόγησης.

Καταγράφοντας τις αρχικές τιμές από τουλάχιστον δύο βαθμολογητές και συσχετίζοντας τις βαθμολογίες που δίνουν αυτοί στις απαντήσεις των εξεταζομένων, μπορούμε να βρούμε το βαθμό συμφωνίας αυτών των βαθμολογητών, υπολογίζοντας και πάλι το δείκτη συνάφειας *Pearson r* ή άλλους δείκτες συνάφειας, εάν οι μετρήσεις δεν είναι αριθμητικές (π.χ. τον δείκτη συνάφειας *Kendall W*, εφόσον οι μετρήσεις είναι τακτικές τιμές). Η μεταξύ των βαθμολογητών αξιοπιστία αναφέρεται στην παράμετρο σφάλματος που προκύπτει λόγω διαφορετικών κριτηρίων βαθμολόγησης και, κατά συνέπεια, χρησιμεύει περισσότερο στο να μας επιτρέψει να

χρησιμοποιήσουμε, σε περαιτέρω διαδικασίες ανάλυσης, τα στοιχεία που έχουμε στα χέρια μας, δηλαδή να τα εμπιστευτούμε στις περαιτέρω διαδικασίες. Δεν μας εξασφαλίζει επομένως ότι το τεστ που χρησιμοποιούμε είναι αξιόπιστο από όλες τις απόψεις, αλλά μας δίνει την ευκαιρία να προχωρήσουμε στη διερεύνηση του θέματος αυτού μεταξύ άλλων. Στην προσπάθεια αυτή συμβάλλουν και άλλοι δείκτες που αναφέρονται στην ομοιογένεια των μετρήσεων μεταξύ των βαθμολογητών (π.χ., *intra-class correlation coefficients*).

Μέρος 4. Εγκυρότητα μετρήσεων και εγκυρότητα των τεστ

Γενικές έννοιες

Όπως και η αξιοπιστία, ως έννοια, έτσι και η εγκυρότητα αναφέρεται και στις μετρήσεις γενικά (ερευνητικές-πειραματικές) αλλά και στα ίδια τα ψυχομετρικά τεστ και αποτελεί καίρια παράμετρο της ευρύτερης επιστημονικής αποδοχής ενός νέου ψυχομετρικού εργαλείου και της θεωρητικής βάσης που το στηρίζει αλλά και στηρίζεται από αυτό.

Αν θελήσω να μετρήσω την τάση του ηλεκτρικού ρεύματος και δεν χρησιμοποιήσω βολτόμετρο, αλλά βάλω το χέρι στην πρίζα, τότε δεν θα μετρήσω την τάση, αλλά τα δευτερόλεπτα ζωής που μου απομένουν. Αν προσπαθήσω να μετρήσω το μήκος του δωματίου και χρησιμοποιήσω, αντί για μέτρο, το μάτι, μάλλον δεν θα μπορέσω να καταλήξω στο σωστό αποτέλεσμα. Οι κατάλληλες κλίμακες μέτρησης και η διαδικασία, γενικότερα, που επιλέγω για τη μέτρηση αυτή, είναι το πρώτο βήμα για την εξασφάλιση εγκυρότητας. Αν θέλω να μετρήσω την ηλικία κάποιου, θα τον ρωτήσω την ημερομηνία γέννησής του και θα υπολογίσω την ακριβή του ηλικία· δεν θα ρωτήσω μόνον αν έχει μικρότερα αδέρφια, ούτε θα ρωτήσω μόνον αν οι γονείς του ζουν. Οι απαντήσεις που λαμβάνουμε πρέπει να είναι, όσο το δυνατόν, διεξοδικότερα πληροφοριακές και για τον λόγο αυτόν, πάντα επιλέγουμε την κατάλληλη κλίμακα μέτρησης. Αυτό, όμως, δεν αρκεί. Χρειάζεται, ανά πάσα στιγμή, να γνωρίζουμε επακριβώς τι θέλουμε να μετρήσουμε, ώστε να μην καταλήξουμε να μετρούμε τελείως διαφορετικές έννοιες από αυτές που μας ενδιαφέρουν. Εάν ένα τεστ νοημοσύνης έχει δημιουργηθεί, με τέτοιον τρόπο, ώστε

τελικά να μετρά, έστω, ικανότητα επεξεργασίας πληροφοριών, τότε δεν είναι τεστ νοημοσύνης. Ένα ακόμη παράδειγμα είναι, όταν με ενδιαφέρει το έκδηλο άγχος και, για να το μετρήσω, κατασκευάζω μια ψυχομετρική κλίμακα με φοβικά ερεθίσματα, για να προκαλέσω την εκδήλωση του άγχους. Στην ουσία είναι πιθανό να καταλήξω να μετρώ τον ουδό ανοχής (*phobic threshold, tolerance*) σε τέτοιου είδους ερεθίσματα και όχι το άγχος που προκαλείται από αυτά.

Με λίγα λόγια, όταν ένα ψυχομετρικό εργαλείο αναφέρεται σε μία συγκεκριμένη έννοια, θα πρέπει να είμαστε σε θέση να αποδείξουμε ότι το τεστ αυτό μετρά τη συγκεκριμένη αυτή έννοια-μεταβλητή και όχι κάτι άλλο. Αυτή η αποδεικτική μας ικανότητα αναφέρεται στην τεκμηρίωση της εγκυρότητας του τεστ.

Υπάρχουν και χρησιμοποιούνται διάφορα είδη εγκυρότητας, εκ των οποίων άλλα καταλήγουν σε αριθμητικούς δείκτες εγκυρότητας και άλλα όχι. Αυτό δεν σημαίνει ότι δεν είναι ισχυρές όσες μέθοδοι ελέγχου της εγκυρότητας δεν καταλήγουν σε αριθμητική έκφραση της εγκυρότητας αυτής. Απεναντίας, κάποιες μέθοδοι που καταλήγουν σε αριθμητικές εκφράσεις της εγκυρότητας δεν είναι τόσο ισχυρές όσο άλλες. Οι κυριότερες μέθοδοι ελέγχου της εγκυρότητας των ψυχομετρικών εργαλείων (οι ίδιες μέθοδοι, με μικρές παραλλαγές, χρησιμοποιούνται και για τον έλεγχο της εγκυρότητας των μετρήσεων γενικά) θα αναλυθούν σύντομα στα επόμενα κεφάλαια.

Κατά τεκμήριο εγκυρότητα

Εάν το τεστ που χρησιμοποιούμε φαίνεται, λογικά, να μετρά αυτό που θέλουμε να μετρά, τότε έχουμε κατά τεκμήριο εγκυρότητα του τεστ. Αυτό το είδος εγκυρότητας είναι το πιο απλό και, θεωρητικά, το πλέον εφικτό, αλλά αυτό δεν σημαίνει ότι δεν παρουσιάζει μειονεκτήματα.

Κατά πρώτο λόγο, η κατά τεκμήριο εγκυρότητα ενός τεστ δεν έχει αποδεικτική ισχύ και ούτε μπορούμε να θεωρήσουμε ότι μόνο αυτή η εγκυρότητα μας αρκεί για να χρησιμοποιούμε με ήσυχη τη συνείδηση το τεστ αυτό. Πολλές φορές, επίσης, η κατά τεκμήριο εγκυρότητα εξαρτάται από το είδος των μεταβλητών που θέλουμε να προσεγγίσουμε και πολλές φορές η ευλογοφάνεια της κατά τεκμήριο εγκυρότητας αποβαίνει επικίνδυνη για τις μετρήσεις μας.

Ας δούμε κάποια γενικά παραδείγματα: Έστω ότι θέλουμε να μετρήσουμε την ικανότητα των υπό εκπαίδευση πιλότων ως προς την παράλληλη, πολλαπλή αντιμετώπιση πληθώρας ερεθισμάτων. Στην περίπτωση, μας ενδιαφέρει να έχουμε αξιόπιστα αποτελέσματα, για να γνωρίζουμε ότι οι πιλότοι αυτοί δεν πρόκειται να παρουσιάσουν άλλη συμπεριφορά από αυτήν που θα προκύψει κατά τη μέτρηση, αλλά ταυτόχρονα θέλουμε να είμαστε και βέβαιοι ότι μετρούμε πραγματικά την ικανότητά τους αυτή και όχι, έστω, τη μηχανική τους ικανότητα (να ενεργοποιούν γρήγορα κάποιους διακόπτες). Δημιουργούμε μια πειραματική διαδικασία, η οποία μπορεί να εφαρμόζεται μέσα σε ένα εξομοιωτή πτήσης και, ζητούμε από τους πιλότους να αντιδράσουν σωστά και άμεσα στις πολύπλοκες απαιτήσεις μιας κατάστασης κρίσης (π.χ., τυφλή προσγείωση). Οποιοσδήποτε συνάδελφος ερευνητής (και δεν είναι απαραίτητο να είναι ούτε συνάδελφος ούτε ερευνητής) πληροφορηθεί για το πείραμα αυτό, θα αντιληφθεί αμέσως ότι πρόκειται για πείραμα ελέγχου των ικανοτήτων των πιλότων.

Το γεγονός ότι το αντικείμενο της ψυχομετρικής διαδικασίας γίνεται αμέσως αντιληπτό, προσδίδει στη διαδικασία κατά τεκμήριο εγκυρότητα. Εάν, από την άλλη, καλέσω τους πιλότους στο εργαστήριο και τους ζητήσω να καταγράψουν τους στίχους ενός τραγουδιού που ακούγεται, ενώ, ταυτόχρονα, εξηγούν σε κάποιον τρίτο τους κανόνες της αεροδυναμικής, θα είναι δύσκολο να καταλάβει κάποιος για ποιο λόγο γίνεται η διαδικασία αυτή. Στο πρώτο παράδειγμα, η ευλογοφάνεια της ψυχομετρικής διαδικασίας δεν θα προκαλέσει αντιδράσεις, τουλάχιστον από την πλευρά των συμμετεχόντων στο δείγμα. Στο δεύτερο παράδειγμα, οι πρώτοι που θα αντιδράσουν είναι οι ίδιοι οι πιλότοι, οι οποίοι δεν θα μπορούν να αντιληφθούν τι σχέση έχει η πειραματική διαδικασία με τις πτήσεις.

Όταν μια ψυχομετρική διαδικασία δεν φαίνεται να έχει κατά τεκμήριο εγκυρότητα, δεν σημαίνει ότι ο ερευνητής που τη δημιούργησε δεν μπορεί να αποδείξει ότι τελικά η διαδικασία είναι κατά τεκμήριο έγκυρη. Συνήθως, εσκεμμένα, αποκρύπτουμε τους πραγματικούς σκοπούς μας, μέσω μιας φαινομενικά άσχετης ψυχομετρικής διαδικασίας. Η μεθοδολογική αυτή προσέγγιση λέγεται παραπλάνηση (*deception*) και εφαρμόζεται πολύ συχνά για να αποφύγουμε πολώσεις που προκαλούνται από επιδράσεις συμμετεχόντων (*participant effects*). Πολλές φορές, όσο πιο ευλογοφανής είναι μια ψυχομετρική διαδικασία, τόσο περισσότερο είναι

αντιληπτή στον συμμετέχοντα ο οποίος, λόγω κοινωνικής αποδοχής, προσπαθεί να δώσει απαντήσεις που θα μας βοηθήσουν να αποδείξουμε αυτό που αντιλαμβάνεται ότι θέλουμε να αποδείξουμε. Φυσικά, ο συμμετέχων μπορεί να αντιδράσει και αντίθετα με αυτά που θέλει ο ερευνητής να καταδείξει, αλλά ο ερευνητής δεν ξέρει ποια είναι η αντίδραση του συμμετέχοντα και *δεν μπορεί να την ελέγξει*. Η επίδραση του κοινωνικά επιθυμητού δεν προκαλεί προβλήματα στα τεστ νοημοσύνης και στα τεστ ικανότητας, επειδή στις περιπτώσεις αυτές, ο συμμετέχων, εκ των πραγμάτων, προσπαθεί να δείξει τον καλύτερό του εαυτό και επομένως, όποιες και αν είναι οι επιδράσεις, είναι μονόδρομες. Παρά το γεγονός αυτό, και πάλι θα πρέπει να γνωρίζουμε ότι η επίδραση αυτή υπάρχει σε διάφορες διαβαθμίσεις, ανάλογα με το πόσο επιθυμητός κοινωνικά θέλει να είναι ο κάθε συμμετέχων.

Κατά συνέπεια, θα πρέπει να γνωρίζουμε ότι σε πολλές περιπτώσεις, η κατά τεκμήριο εγκυρότητα, όπως γίνεται αντιληπτή από τους συμμετέχοντες, ίσως και να μην είναι επιθυμητή. Από την άλλη, είναι πάρα πολύ δύσκολο να βρει κανείς μια ψυχομετρική, μια πειραματική, ή μια ερευνητική, γενικότερα, διαδικασία που να μην διαθέτει κατά τεκμήριο εγκυρότητα, καθώς όλες αυτές οι διαδικασίες σχεδόν πάντα βασίζονται σε κάποια γνωστά θεωρητικά πλαίσια και τεκμηριώνονται μέσω αυτών. Ίσως οι μόνες περιπτώσεις που δεν διαθέτουμε κατά τεκμήριο εγκυρότητα είναι αυτές που η αντίστοιχη ψυχομετρική διαδικασία αποτελεί την απαρχή μιας νέας επιστημονικής θεώρησης. Επίσης, υπάρχουν και διαβαθμίσεις στην κατά τεκμήριο εγκυρότητα. Άλλη είναι η κατά τεκμήριο εγκυρότητα που διαθέτουν τα τεστ ικανοτήτων και άλλη αυτή που διαθέτουν τα τεστ αξιών.

Εγκυρότητα αντιπροσωπευτικού περιεχομένου

Ένα από τα πρώτα ερωτήματα που τίθενται στους ερευνητές-δημιουργούς ψυχομετρικών εργαλείων είναι: «ελέγχει όλες τις πλευρές της μηχανικής ικανότητας, ή σου έχουν ξεφύγει μερικές;» Πολλές φορές ο δημιουργός του τεστ δεν είναι σε θέση να απαντήσει και αυτό συμβαίνει γιατί δεν είναι βέβαιος για την εγκυρότητα αντιπροσωπευτικού περιεχομένου του τεστ. Όταν ο *G. W. Allport* υποστήριξε την αναλυτική περιγραφή της προσωπικότητας, θέλησε να απαριθμήσει έναν κατάλογο γνωρισμάτων προσωπικότητας και θέλησε να είναι ο κατάλογος αυτός, όσο το δυνατόν, πλήρης, επειδή πίστευε ότι οι τυπολογίες απομονώνουν ένα

μόνο τμήμα της ανθρώπινης προσωπικότητας, την οποία, πλέον, ήταν ανάγκη οι ψυχολόγοι να αντιμετωπίσουν στην ολότητά της. Η αρχή έγινε με την αναζήτηση τέτοιων χαρακτηριστικών προσωπικότητας σε όλα τα λεξικά της αγγλικής γλώσσας και την καταγραφή 18.000 περίπου λέξεων, από τις οποίες τελικά επελέγησαν περίπου 4.500 επίθετα τα οποία ονόμασε *traits*, χαρακτηριστικά προσωπικότητας και τα ταξινόμησε με βάση τη σημασία τους για την προσωπικότητα.

Ο *M. Rokeach* από την άλλη, θέλοντας να κατασκευάσει ένα τεστ μέτρησης των αξιών, συγκέντρωσε 18 αξίες-σκοπούς (*terminal values*) και 18 αξίες-μέσα (*instrumental values*), τις οποίες και κατέταξε σε δύο καταλόγους, προς ιεράρχηση, ερωτημάτων. Ο ίδιος, όμως, τονίζει ότι άλλοι ερευνητές μπορούν να καταλήξουν σε τελείως διαφορετικούς καταλόγους αξιών και ότι η επιλογή των 36 αυτών αξιών ήταν κατά κύριο λόγο «διαισθητική»¹¹. Σε αντίθεση με την διαισθητική αυτήν πρακτική δημιουργίας ενός τεστ, ο *Παρασκευόπουλος*, θέλοντας να δημιουργήσει μια ψυχομετρική κλίμακα μέτρησης της προβληματικής συμπεριφοράς, προσπάθησε να συγκεντρώσει τις διάφορες εκφάνσεις της συμπεριφοράς αυτής σε παιδιά, μέσω αρχικών συνεντεύξεων με γονείς, δασκάλους, υπευθύνους βρεφονηπιακών σταθμών, εργαζόμενους σε νοσοκομεία για παιδιά, κ.λπ., δηλαδή, προσπάθησε *εξ αρχής* να καταγράψει πλήρως όλο το εύρος της προβληματικής συμπεριφοράς *όπως αυτή παρατηρείται στην πραγματικότητα*. Κατόπιν, προχώρησε σε επιλογή των πλέον αντιπροσωπευτικών από αυτά τα συμπτώματα προβληματικής συμπεριφοράς, αφού τα δοκίμασε σε διαδοχικές έρευνες πιλότους, για να καταλήξει σε ένα κατάλογο 67 χαρακτηριστικών προβληματικών συμπεριφορών, τον οποίο και συνέταξε σε ερωτηματολόγιο που συμπληρώνεται από το δάσκαλο ή τον γονέα και καταλήγει σε τέσσερις παράγοντες προβληματικής συμπεριφοράς: αναστολή-νεύρωση, ανεπάρκεια-ανωριμότητα, αντικοινωνική συμπεριφορά και, ψυχοσωματική συνδρομή¹². Με την ίδια λογική, εάν θελήσουμε να συντάξουμε ένα ψυχομετρικό εργαλείο το οποίο θα μετρά την εκπαίδευση των ψυχολόγων στην Ελλάδα, η πρώτη μας κίνηση θα είναι να καταστρώσουμε ένα προσχέδιο, μέσω ανασκόπησης της βιβλιογραφίας, και να προσεγγίσουμε επαγγελματίες ψυχολόγους καθώς και πανεπιστημιακούς δασκάλους και φοιτητές για να πληροφορηθούμε για όσο το

¹¹ Βλέπε σύγγραμμα: **M. Rokeach** (1973) *The Nature of Human Values*. The Free Press. Σελ. 30.

δυνατόν περισσότερες πλευρές επί του θέματος, ώστε μετά να μπορέσουμε να προχωρήσουμε με πληρότητα. Εννοείται, ότι, για να μπορέσουμε να κάνουμε αυτές τις διαδικασίες, θα πρέπει να προσφέρεται και το αντικείμενο της όλης ψυχομετρικής προσπάθειας. Το να ανακαλύψουμε, για παράδειγμα, όλες ταυτόχρονα τις πραγματικές πτυχές της γνωστικής επεξεργασίας πληροφοριών είναι ένα τεράστιο εγχείρημα και το πιθανότερο είναι πως θα αποτύχουμε οικτρά.

Πολλοί θεωρούν την εγκυρότητα αντιπροσωπευτικού περιεχομένου ως μια «βελτιωμένη έκδοση» της κατά τεκμήριο εγκυρότητας. Αυτή η ταύτιση δεν ισχύει, καθώς η αντιπροσωπευτικού περιεχομένου εγκυρότητα αλληλεπικαλύπτεται με την κατά τεκμήριο εγκυρότητα, αλλά ταυτόχρονα, αποτελεί το αποδεικτικό τμήμα της τελευταίας. Είναι, κατά τον τρόπο αυτόν, η εγκυρότητα αντιπροσωπευτικού περιεχομένου, το μη-εύλογο τμήμα της κατά τεκμήριο εγκυρότητας και για τον λόγο αυτό, τις περισσότερες φορές, για να έχουμε αντιπροσωπευτικού περιεχομένου εγκυρότητα, θα πρέπει πρώτα να έχουμε εγκυρότητα κατά τεκμήριο. Στη διεθνή βιβλιογραφία, η αντιπροσωπευτικού περιεχομένου εγκυρότητα αναφέρεται ως *criterion-related validity* (ως προς το κριτήριο εγκυρότητα), καθώς είναι πολλές οι περιπτώσεις που υπάρχουν αντικειμενικά κριτήρια με βάση τα οποία καταρτίζονται, εξ αρχής, τα διάφορα ψυχομετρικά εργαλεία. Για παράδειγμα, εάν θέλουμε να μετρήσουμε τη μαθηματική ικανότητα παιδιών και εφήβων, δεν θα συμπεριλάβουμε και πανεπιστημιακές γνώσεις στο ψυχομετρικό μας εργαλείο, μόνο και μόνο επειδή αποτελούν τμήμα της μαθηματικής επιστήμης, αλλά θα χρησιμοποιήσουμε ερωτήματα και προβλήματα που αναφέρονται σε όσα έχουν διδαχθεί οι μαθητές μέχρι και την Γ' λυκείου. Στην περίπτωση, η «επιλογή» αυτή οφείλεται στο είδος του στατιστικού πληθυσμού που θέλουμε να προσεγγίσουμε, αλλά, υπάρχουν και άλλες περιπτώσεις όπου το ίδιο το αντικείμενο μέτρησης ορίζει το πλαίσιο ευρύτητας των επιλεγμένων ψυχομετρικών ερεθισμάτων.

Κατά κόρον, η εγκυρότητα αντιπροσωπευτικού περιεχομένου αποτελεί τον ακρογωνιαίο λίθο υποστήριξης της εγκυρότητας των διαφόρων τεστ. Δεν εκφράζεται αριθμητικά και δεν θα έπρεπε· το είδος αυτό της εγκυρότητας είναι τόσο ευαίσθητο και στην παραμικρή αλλαγή, ώστε το να έχει κανείς στο τεστ που έχει δημιουργήσει

¹² Βλέπε σύγγραμμα: **I.N. Παρασκευόπουλος** (1992) *Ψυχολογία Ατομικών Διαφορών*. Σελ. 82-85.

90% αντιπροσωπευτικού περιεχομένου εγκυρότητα δεν έχει νόημα, καθώς το υπόλοιπο 10% μπορεί να είναι ιδιαίτερο σημαντικό και να έχει αγνοηθεί. Ή έχουμε, λοιπόν, ή δεν έχουμε εγκυρότητα αντιπροσωπευτικού περιεχομένου, είτε, δηλαδή, καταφέραμε ως ψυχομέτρες να αντιπροσωπεύσουμε όλο το εύρος, είτε όχι. Οι ενδιάμεσες καταστάσεις είναι παρακινδυνευμένες.

Εγκυρότητα προβλεπτικής συνάφειας ή προβλεπτική εγκυρότητα

Ένα πολύ ισχυρό όπλο που διαθέτουμε στην ψυχολογία είναι αυτό της πρόβλεψης. Εξάλλου, ο στόχος και της θεωρητικής αλλά και της εφαρμοσμένης ψυχολογίας είναι ακριβώς αυτός, δηλαδή, να μπορέσει να προβλέψει τις μελλοντικές εξελίξεις ως προς τις ψυχολογικές παραμέτρους του κάθε ατόμου, με βάση την τωρινή πραγματικότητα. Για τον λόγο αυτό, η εγκυρότητα προβλεπτικής συνάφειας είναι η ισχυρότερη από στατιστικής απόψεως.

Για να επιτύχουμε εγκυρότητα προβλεπτικής συνάφειας, θα πρέπει να διαθέτουμε, εκτός από το ψυχομετρικό εργαλείο που θα χρησιμοποιήσουμε, και μια αντικειμενική μέτρηση της ίδιας έννοιας, σε κάποιο σημείο του μέλλοντος. Έτσι, εάν ένα ψυχομετρικό εργαλείο «προβλέψει» την είσοδο κάποιου στο πανεπιστήμιο, αυτό το μέσο θα πάψει να είναι «κρυστάλλινη σφαίρα» όταν αυτός πετύχει στις εισαγωγικές εξετάσεις του πανεπιστημίου. Εάν μάλιστα μια τέτοια προβλεπτική ικανότητα αποδειχθεί για δείγμα ατόμων τα οποία μετρήθηκαν ως προς τις «ακαδημαϊκές» τους ικανότητες, και αφού βρέθηκαν να έχουν υψηλές τέτοιες ικανότητες, πέρασαν μετά από 6 μήνες στο πανεπιστήμιο, τότε λέμε ότι το τεστ αυτό διαθέτει προβλεπτική εγκυρότητα και συναγωνιζόμαστε επαξίως (και επικερδώς), μέσω των τηλεφωνικών καλωδίων, τις αστρολογικές προβλέψεις.

Για να μπορέσουμε βέβαια να ελέγξουμε το αν οι προβλέψεις μας ήταν σωστές, θα πρέπει να εξακολουθήσουμε να έχουμε πρόσβαση στους αρχικούς συμμετέχοντες στο δείγμα μας και μάλιστα, κατ' αντιστοιχία, ώστε, για κάθε έναν, να έχουμε την αρχική ψυχομετρική πρόβλεψη και την τελική απόληξη. Έχοντας τα στοιχεία αυτά στα χέρια μας, μπορούμε να υπολογίσουμε τη συνάφεια που προκύπτει και να την ερμηνεύσουμε ανάλογα. Το γεγονός βέβαια, ότι αναφερόμαστε σε σύνολα ατόμων και συνάφεια μεταξύ τωρινών και μελλοντικών μετρήσεων, δεν θα πρέπει να

αφήνει περιθώρια αμφιβολίας ότι, για μια ακόμη φορά, αναφερόμαστε σε δειγματοληπτικές διαδικασίες και κατανομές. Υπάρχουν, όμως, και προβλήματα.

Το σοβαρότερο είναι ότι το ψυχομετρικό μας εργαλείο, εκτός από επιτυχία, προβλέπει και αποτυχία. Στην περίπτωση του παραπάνω παραδείγματος, είναι εύκολο να δούμε αν κάποιος πέτυχε ή απέτυχε στο πανεπιστήμιο και έτσι το πρόβλημα δεν είναι ιδιαίτερο. Ας σκεφτούμε όμως μια διαφορετική περίπτωση: Διαθέτουμε ένα τεστ ταχύτητας και ακρίβειας το οποίο και χορηγούμε σε 200 διαγωνιζόμενους για 50 τραπεζικές θέσεις εργασίας. Οι 50 ταχύτεροι και ακριβέστεροι επιλέγονται από την τράπεζα, ως υπάλληλοί της και αρχίζουν να εργάζονται. Το ζήτημα είναι, αν το τεστ που χρησιμοποιήσαμε και το οποίο θεωρητικά προβλέπει ότι οι συγκεκριμένοι υπάλληλοι θα είναι ταχείς και ακριβείς, δικαιώνεται ή όχι. Αναζητούμε, λοιπόν, τους 50 επιλεγέντες για να δούμε «πόσο αποδίδουν» στην εργασία τους. Ανεξαρτήτως της επίδοσης αυτών των 50 προσληφθέντων υπαλλήλων, παραμένει ένας αριθμός 150 διαγωνισθέντων, για τους οποίους ποτέ δεν θα μάθουμε πώς θα τα πήγαιναν στη εργασία αυτή, αν είχαν επιλεγεί. Μειωμένα στοιχεία λοιπόν και το πρόβλημα αυτό συνήθως «λύνεται» διαισθητικά και μόνο, αλλά στην ουσία, μπορούμε να αναφερθούμε μόνο σε εκείνα τα στοιχεία τα οποία έχουμε στα χέρια μας και *σε τίποτε περαιτέρω*.

Όπου είναι δυνατόν να χρησιμοποιηθεί η εγκυρότητα προβλεπτικής συνάφειας, καλό θα είναι να προσπαθούμε να την υπολογίσουμε, παρόλα τα πρακτικά και στατιστικά προβλήματα που μπορεί να μας παρουσιάσει. Εάν καταφέρουμε να προβλέψουμε τη μελλοντική επίδοση, συμπεριφορά, ασθένεια, έστω και για τμήμα των πιθανών εκβάσεων της ψυχομετρικής διαδικασίας, αυτό *είναι επίτευγμα*.

Εγκυρότητα συγχρονικής συνάφειας

Το είδος αυτό εγκυρότητας είναι μεθοδολογική εφεύρεση. Εξυπηρετεί κυρίως τις ερευνητικές ανάγκες, καθώς στο πεδίο αυτό, τα ερωτήματα ενδιαφέροντος είναι εστιασμένα και πολύ εξειδικευμένα. Χρησιμοποιείται, όμως, ευρύτατα και από την ψυχομετρία, επειδή αποτελεί ισχυρότατο αποδεικτικό στοιχείο της αρτιότητας ενός ψυχομετρικού τεστ.

Όταν κάποιος δημιουργήσει ένα νέο τεστ, αναζητά προηγούμενα τεστ της ίδιας «κατηγορίας», τεστ δηλαδή που αναφέρονται στην ίδια μεταβλητή-έννοια και καταλήγουν σε παρόμοιου τύπου μετρήσεις. Τα προηγούμενα αυτά τεστ βασίζονται σε κάποιο θεωρητικό υπόβαθρο. Το νέο τεστ δεν είναι απαραίτητο να βασίζεται στα ίδια ή έστω παρόμοια πλαίσια, καθώς είναι δυνατόν και ταυτόχρονα θεμιτό να στηριχθεί από (και, αν τα καταφέρει, να στηρίξει) ένα καινούριο θεωρητικό πλαίσιο. Ο στόχος, βέβαια, είναι η χορήγηση και του παλαιότερου αλλά και του νέου τεστ μέσω δειγματοληπτικής διαδικασίας και η σύγκριση-συσχέτιση των δύο τελικών μετρήσεων για την ίδια μεταβλητή.

Με τη λογική αυτή, ο *Bennet* κατασκεύασε το 1962 το *Differential Aptitude Test* (DAT) και, χορηγώντας το, χορήγησε εν παραλλήλω το *Primary Mental Ability* τεστ (PMA, *Thurstone & Thurstone*, 1962) και αργότερα, το *General Aptitude Test Battery* (GATB, *The USA Government*, 1970 - όπως είναι προφανές, η διαδικασία ελέγχου της εγκυρότητας έχει μεγάλη διάρκεια). Οι δείκτες εγκυρότητας συγχρονικής συνάφειας ήταν ιδιαίτερα υψηλοί, επιβεβαιώνοντας την επιτυχία μέτρησης των ικανοτήτων μέσω του DAT. Όπως, όμως, αναφέρει ο *Kline*¹³, το DAT υστερεί στον τομέα τη διαφοροποιητικής ισχύος (βλέπε και Κεφάλαιο 'Αξιοπιστία των ημικλάσεων')· επομένως, ίσως και το PMA παρουσιάζει το ίδιο μειονέκτημα. Εκτός, λοιπόν, από το πρόβλημα που αναφέρεται στην ανεύρεση των καταλλήλων, προς συσχέτιση, τεστ, τίθεται το ερώτημα «μήπως και τα δύο τεστ είναι ανέγκυρα;» Η απάντηση μπορεί να δοθεί μόνο μέσα από τη λεπτομερή συγκέντρωση πληροφοριών για το ποια τεστ παρουσιάζουν συγχρονική συνάφεια με ποια και πόσα, και για το ποια είναι εκείνα που φαίνεται να μη μετέχουν στο δίκτυο αυτό πληροφοριών.

Ένα άλλο ζήτημα, είναι αυτό της διπλής χορήγησης των τεστ στους συμμετέχοντες του δείγματος. Η διπλή αυτή χορήγηση μπορεί να επηρεάζει τα αποτελέσματα του ενός από τα δύο τεστ, καθώς χρονικά το ένα ακολουθεί το άλλο, κατά τη χορήγηση. Η μοναδική λύση είναι να χορηγηθούν τα τεστ με την αντίστροφη σειρά στους μισούς συμμετέχοντες, αλλά και πάλι δεν μπορούμε να είμαστε απόλυτα βέβαιοι για το μέγεθος της επίδρασης του ενός τεστ στο άλλο.

¹³ **P. Kline** (1994) *The Handbook of Psychological Testing*. Routledge. Σελ. 428.

Υπάρχουν και περιπτώσεις, όπου ο τελικός δείκτης συγχρονικής συνάφειας δεν είναι απλά υψηλός, αλλά πολύ, έως υπερβολικά, υψηλός. Στην περίπτωση, τα δύο τεστ φαίνεται να μετρούν τα ίδια πράγματα, και συνεπώς, το ένα από τα δύο θα πρέπει να εκλείψει. Στο σημείο αυτό, προκύπτει το ερώτημα του τι καινούριο προσφέρει το νέο τεστ, ερώτημα που τίθεται ούτως ή άλλως. Εδώ, καλείται ο δημιουργός του να παραθέσει λόγους για τους οποίους το δικό του τεστ θα πρέπει να χρησιμοποιείται στη θέση άλλων, παλαιότερων τεστ. Οι λόγοι αυτοί συνήθως είναι πρακτικού τύπου (απαιτείται λιγότερο χρονοβόρα χορήγηση, το τεστ δεν χρειάζεται εξειδικευμένους συνεργάτες για να χορηγηθεί, τα ερωτήματα είναι καλύτερα διατυπωμένα, περιλαμβάνουν περισσότερους τομείς ενδιαφέροντος, κ.τ.λ.), αλλά θα πρέπει να λαμβάνουμε υπόψιν και το τι θεωρητικά νέο έχει να προσφέρει το τεστ αυτό. Δηλαδή, θα πρέπει να ελέγξουμε το αν κάποια μελανά σημεία της ισχύουσας θεώρησης στον αντίστοιχο χώρο μπορούν, μέσω του νέου αυτού τεστ, να διευκρινισθούν. Εάν κάτι τέτοιο μπορεί να αποδειχθεί, τότε, σαφέστατα, το νέο αυτό τεστ θα πρέπει να αντιμετωπισθεί ως ικανοποιητική εναλλακτική λύση.

Εγκυρότητα εννοιολογικής δομής

Αυτό το είδος εγκυρότητας αναφέρεται στην *de facto* απόδειξη της ικανότητάς του να καταλήγει στις μετρήσεις του τύπου που ενδιαφέρουν. Η λογική της εγκυρότητας εννοιολογικής δομής στηρίζεται στο ότι υπάρχουν ομάδες ανθρώπων, οι οποίες εκφράζουν, μέσω των χαρακτηριστικών τους, διάφορες διαβαθμίσεις των θεωρητικών μετρήσεων του κάθε τεστ. Έτσι, αν ένα τεστ αναφέρεται στη μαθηματική ικανότητα, τότε θα περιμένουμε από δείγμα μαθηματικών υψηλές επιδόσεις στο τεστ αυτό, ενώ από δείγμα φιλολόγων χαμηλές επιδόσεις. Αν ένα τεστ μετρά αξίες, αναμένουμε υψηλή αισθητική αξία από φιλολόγους και υψηλή θρησκευτική αξία από θεολόγους. Είτε μας δυσαρεστεί είτε όχι, υπάρχουν «ετικέττες» που μας συνοδεύουν και πολλές φορές προκαθορίζουν τη συμπεριφορά και τη διαμόρφωση στάσεων και αξιών. Εάν, επομένως, ένα τέτοιο τεστ αξιών κατορθώσει για κάποιες επαγγελματικές ή κοινωνικές ομάδες να καταδείξει ότι παρουσιάζουν τις σύμφωνα με τη θεωρία αναμενόμενες αξίες, τότε το τεστ αυτό διαθέτει εγκυρότητα εννοιολογικής δομής.

Η διαδικασία, όμως, που ακολουθείται έχει περαιτέρω, βαθύτερους στόχους. Ο βασικότερος είναι η δημιουργία ενός πυκνού δικτύου πληροφοριών γύρω από τις υπό μελέτη μεταβλητές και θεωρητικά πλαίσια. Η ψυχομετρική διαδικασία αποτελεί, όταν ξεκινήσει η συλλογή των πληροφοριών για το δίκτυο αυτό, μέσο και έναυσμα, ενώ παύει να είναι ο κυρίαρχος στόχος. Γύρω από την υπό μελέτη έννοια συσσωρεύονται πολλές άλλες έννοιες και μεταβλητές που μπορεί να προέρχονται από τον ίδιο ή και διαφορετικό θεωρητικό χώρο. Ταυτόχρονα, συσσωρεύονται και στοιχεία που αφορούν στις ειδικές συνθήκες που επικρατούν κατά την ψυχομετρική διαδικασία, ή δρουν αυτόνομα, και διαμορφώνουν τις μετρήσεις που προκύπτουν από το τεστ. Η συγκέντρωση αυτή των πληροφοριών οδηγεί πρώτιστα στην οργάνωση του δικτύου στο οποίο καταχωρούνται οι πληροφορίες αυτές. Το δίκτυο αυτό δεν είναι «ιδιοκτησία» του δημιουργού του συγκεκριμένου τεστ, αλλά το χειρίζονται και το διαμορφώνουν όλοι όσοι θεωρητικοί, ερευνητές και επαγγελματίες ασχολούνται με τα αντίστοιχα θεωρητικά ή εφαρμοσμένα πλαίσια. Ο κάθε ειδικός, θα προσθέσει, μέσω ερευνητικής διαδικασίας ή εφαρμογής των ερευνητικών πορισμάτων ή μέσω της πλήρους περιγραφής των μέχρι στιγμής συσσωρευμένων πληροφοριών, νέες πληροφορίες στο δίκτυο αυτό. Όσες περισσότερες πληροφορίες διαθέτει το δίκτυο αυτό σχετικά με το συγκεκριμένο θέμα, που εθίγη όταν ο δημιουργός ενός τεστ θέλησε να αποδείξει την εγκυρότητα του τεστ, τόσο περισσότερο και η εγκυρότητα αυτή αποδεικνύεται αλλά και τόσο περισσότερο αναπτύσσονται νέες προσπάθειες, είτε ψυχομετρικές είτε άλλου είδους.

Η διαδικασία της αναζήτησης της εγκυρότητας εννοιολογικής δομής παραλληλίζεται συνήθως με τη φιλοσοφική θεώρηση της επιστήμης του *K. Popper*. Δεν θα είμαστε ποτέ σε θέση να αποδείξουμε την εγκυρότητα εννοιολογικής δομής κάποιας ψυχομετρικής κλίμακας, αλλά τα αποδεικτικά στοιχεία θα συσσωρεύονται (*corroborated evidence*) στο δίκτυο πληροφοριών και θα προσεγγίζουμε συνεχώς την εγκυρότητα που αναζητούμε. Πώς άλλωστε να μην παραλληλίσουμε τη θεωρία της εγκυρότητας με αυτή του *Popper*, όταν και οι δύο αναφέρονται στην αναζήτηση δύο μορφών αλήθειας; Καθώς όμως, η διαρκής αναζήτηση της εγκυρότητας προϋποθέτει συνεχή τροφοδότηση του δικτύου πληροφοριών με νέα στοιχεία, είναι αναγκαία και η συνεχής ερευνητική δραστηριότητα που θα αποδώσει τα στοιχεία αυτά, αλλά ταυτόχρονα θα αποτελέσει και νέο σημείο εκκίνησης.

Χαρακτηριστικά, ως παράδειγμα αναζήτησης και επίτευξης εγκυρότητας εννοιολογικής δομής, θα αναφέρουμε το EPQ (*Eysenck Personality Questionnaire*) και θα ασχοληθούμε με την διαδικασία που εφάρμοσε ο *H. Eysenck* για να αποδείξει την εγκυρότητα αυτή. Ως γνωστόν, η θεωρία που ανέπτυξε ο *Eysenck* είναι θεωρία προσωπικότητας και αναφέρεται σε τρεις διαστάσεις: την εξωστρέφεια-εσωστρέφεια, τον νευρωτισμό και τον ψυχωτισμό. Θα πρέπει να προσθέσουμε ότι το τεστ που κατασκεύασε ο *Eysenck* περιλαμβάνει και μια κλίμακα ψεύδους, ώστε να είναι δυνατή η εκτίμηση της κοινωνικής αποδοχής που παρεισφύει και επιδρά στις μετρήσεις. Ο *Eysenck* σκέφτηκε ότι θα μπορούσε να καταδείξει για τα διάφορα χαρακτηριστικά προσωπικότητας που προέκυπταν από το EPQ, αντίστοιχες συμπεριφορές στους κατάλληλους στατιστικούς πληθυσμούς. Εξάλλου, δεν ήταν δυνατό να συσχετισθούν οι μετρήσεις του EPQ με κάποιες αντικειμενικές μετρήσεις εσωστρέφειας-εξωστρέφειας, νευρωτισμού και ψυχωτισμού, και ακόμη και αν συνέβαινε αυτό, τότε ποιο θα ήταν το νόημα του EPQ; Κατάφερε λοιπόν να δείξει σε πολλές περιπτώσεις την αναμενόμενη συμπεριφορά, έστω των εξωστρεφών, για άτομα που στο EPQ είχαν υψηλές μετρήσεις εξωστρέφειας. Για παράδειγμα, οι εξωστρεφείς παρουσιάζουν αργούς ρυθμούς συντελεστικής μάθησης, σε αντίθεση με τους εσωστρεφείς. Καταδεικνύοντας, λοιπόν, μια τέτοια συμπεριφορά για άτομα που σύμφωνα με το EPQ ήταν εξωστρεφείς, είχε διαθέσιμες δύο πληροφορίες τις οποίες και ενέταξε στο δίκτυο πληροφοριών που δημιουργήθηκε γύρω από το EPQ. Αντίστοιχα, υποστήριξε και τις υπόλοιπες κλίμακες του τεστ που είχε δημιουργήσει, αλλά το σημαντικότερο είναι ότι, εγκαθιδρύοντας το δίκτυο πληροφοριών γύρω από την αναζήτηση της εγκυρότητας εννοιολογικής δομής, οδηγήθηκε στην ανάπτυξη και τελειοποίηση της θεωρίας της προσωπικότητας και, ταυτόχρονα, οδήγησε εκατοντάδες ερευνητών στην αναζήτηση μετεξελίξεων της θεωρίας αυτής.

Ένα ακόμη χαρακτηριστικό παράδειγμα, είναι αυτό του MMPI. Πιο κοντά στην κλασική εφαρμογή της θεωρίας της εγκυρότητας εννοιολογικής δομής το MMPI, χορηγήθηκε σε διάφορων ειδών κλινικά δείγματα, όπως αυτά είχαν διαγνωσθεί μέσω των κριτηρίων του DSM και του ICD), καθώς και σε εγκλείστους σωφρονιστικών καταστημάτων, ιδρυματικούς υπερήλικες κ.λπ. και διαπιστώθηκε ότι οι μετρήσεις του MMPI συμφωνούσαν με την «πραγματικότητα» στην κάθε περίπτωση. Φυσικά, οι χορηγήσεις αυτές δεν έγιναν από τη μία μέρα στην άλλη· για

να ολοκληρωθεί μια τέτοια διαδικασία, χρειάζονται πολλά έτη και συνεχείς ερευνητικές προσπάθειες.

Μέρος 5. Πολώσεις

Σύμφωνα με τον Francis Bacon για να παρατηρήσει κανείς κάτι με ακρίβεια, πρέπει να αποφύγει συγκεκριμένα «είδωλα». Αυτά τα είδωλα είναι τεσσάρων τύπων: “*Idols of the Tribe*” (σε ελεύθερη μετάφραση, είδωλα της Φυλής) ή προκαταλήψεις και τάσεις κοινές σε όλους τους ανθρώπους· λόγω των ειδώλων αυτών, που κάθε άνθρωπος συντηρεί στο υποσυνείδητό του, η παρατήρηση χάνει σε αξιοπιστία και καταλήγει σε αποκλίνοντα της αλήθειας συμπεράσματα. Αν οι πολώσεις αυτές είναι σχετικά γνωστές και επομένως μπορούν κατά κάποιον τρόπο να προβλεφθούν και να απαλειφθούν, τα “*Idols of the Cave*” (είδωλα του σπηλαίου) ή οι προκαταλήψεις και οι λανθασμένες εκτιμήσεις του κάθε ανθρώπου λόγω του περιορισμένου εύρους εμπειριών του δεν είναι το ίδιο εύκολο να αντιμετωπισθούν καθώς η εμπειρία και οι γνώσεις του καθενός είναι ίδιον της προσωπικότητας και αντικείμενο μελέτης *per se*. παράλληλα, τα “*Idols of the Forum*” (είδωλα της αγοράς), είδωλα δηλαδή που δημιουργούνται κυρίως λόγω παρερμηνειών ή «γλώσσας» κατά την κοινωνική αλληλεπίδραση επιτείνουν τα προβλήματα πόλωσης· τέλος, τα “*Idols of the Theater*” (είδωλα αυθεντίας), δηλαδή εκείνα τα είδωλα που προέρχονται από την τυφλή αποδοχή φιλοσοφικών, θεολογικών και άλλων συστημάτων που έχουν προταθεί *ex cathedra* ολοκληρώνουν την εικόνα σύγχυσης και κατευθύνουν σε συστηματοποίηση εσφαλμένων νόμων και αποδοχών σε τελολογικό σύστημα. Αυτά, μάλιστα τα θεωρητικά συστήματα, θα πρέπει να τονιστεί, ερμηνεύονται με διαφορετικό τρόπο από κάθε άνθρωπο λόγω και των “*Idols of the Cave*”.

Η πιο πάνω προσέγγιση δεν είχε σκοπό τη Φιλοσοφική ανάλυση των πολώσεων, αλλά αποτελεί ένα απλό παράδειγμα των κινδύνων και παρερμηνειών που ελοχεύουν σε όλη την πορεία της παρατήρησης, της έρευνας και της μέτρησης, ειδικά όταν οι μετρήσεις αυτές αφορούν σε ανθρώπινες ιδιότητες και χαρακτηριστικά που προέρχονται από τις ατομικές διαφορές. Όμως, οι πολώσεις, οι παρερμηνείες και οι επιπτώσεις τους στη μέτρηση των ανθρώπινων χαρακτηριστικών δεν είναι αδιάφορο ψυχομετρικό σφάλμα και πρέπει να αντιμετωπίζονται κατά το στάδιο της δημιουργίας ενός ψυχομετρικού τεστ ή, έστω, να υπάρχει τρόπος εξάλειψης των

πολώσεων κατά τη μέτρηση ή και μετά από αυτήν, αλλά σίγουρα πριν την εξαγωγή τελικών συμπερασμάτων για το άτομο.

Είδη πολώσεων

Συνήθως, στην Ψυχομετρία αντιμετωπίζουμε τρία είδη πολώσεων: πολώσεις ερωτημάτων (*item bias*), εσωτερικές πολώσεις (*intrinsic test bias*) και εξωτερικές πολώσεις του τεστ (*extrinsic test bias*). Η έννοια της πόλωσης και η θεωρητική μελέτη της δεν σημαίνει απαραίτητα ότι θα ληφθεί απόφαση και για την αναπροσαρμογή των πολωμένων τιμών, καθώς υπάρχουν διαφόρων ειδών εξωτερικοί ή κοινωνικοί παράγοντες που δημιουργούν τις πολώσεις αυτές. Σε πολλές περιπτώσεις ίσως μάλιστα, δεν είναι θεμιτή η ανεύρεση των πραγματικών τιμών και η αναπροσαρμογή των φαινομενικών τιμών σύμφωνα με τη θεωρία των αληθών τιμών. Ένα παράδειγμα είναι αυτό των φοιτητών που θα μπορούσαν να επιτύχουν καλύτερη βαθμολογία εάν δεν είχαν κουραστεί υπέρ το δέον διαβάζοντας ολόκληρη την προηγούμενη νύχτα. Γνωρίζουμε δηλαδή, έστω για 10 φοιτητές ότι επέτυχαν τον βαθμό 6 ενώ θα μπορούσαν να γράψουν καλύτερα αν δεν είχαν εξαντλήσει τις πνευματικές τους δυνάμεις. Ακόμη και αν γνωρίζουμε ότι και οι 10 θα έγραφαν κατά μία μονάδα καλύτερα (δηλαδή ότι όλοι θα έγραφαν πραγματικές τιμές “7”), πρέπει να αναπροσαρμόσουμε τις φαινομενικές τους τιμές; Τίθεται επίσης, σε άλλο παράδειγμα, το ζήτημα της ύλης που διαβάζει ένας φοιτητής. Εάν διαβάσει τη μισή ύλη, τότε μια πιθανή αποτυχία του (πόλωση του αποτελέσματος) είναι θέμα δικής του επιλογής και όχι θέμα τύχης, ούτε οφείλεται σε άλλους παράγοντες μόνον. Στην περίπτωση αυτή ο βαθμός του δεν πρέπει να αναπροσαρμοσθεί, με βάση τη θεωρία των αληθών τιμών.

Από την άλλη, εάν ένα τεστ γνώσεων περιλαμβάνει ερωτήσεις με επιλογές «ΝΑΙ» και «ΟΧΙ» τότε ο εξεταζόμενος έχει 50% πιθανότητες να απαντά κάθε ερώτηση σωστά (ή λάθος) στην τύχη. Συνολικά, σε όλες τις ερωτήσεις-δοκιμασίες η πιθανότητα των επιτυχημένων λόγω τύχης απαντήσεων εξακολουθεί να είναι 50% καθώς οι απαντήσεις σε κάθε ερώτημα μπορούν να θεωρηθούν ανεξάρτητα πιθανοθεωρητικά ενδοχόμενα. Αυτή η πραγματικότητα (που ισχύει σε όλα τα τεστ πολλαπλών επιλογών) και η οποία αντικατοπτρίζει μία από τις συνηθέστερες πολώσεις οδηγεί σε βέβαιη αναπροσαρμογή των φαινομενικών τιμών με βάση το

συνολικό αριθμό των λαθών του εξεταζομένου που αποτελεί ένδειξη για την πόλωση λόγω επιδράσεων τύχης. Αυτή η λογική αναπροσαρμογής δεν γίνεται «αποδεκτή» κυρίως από τα παιδιά που θεωρούν ότι η τυχαία απάντηση σε μία ερώτηση αποτελεί και αυτή επιτυχία, εφόσον έχει επιτευχθεί ο στόχος της ανεύρεσης της σωστής απάντησης, έστω και χωρίς να την γνωρίζουν.

Γενικές πολώσεις για κάποιο τεστ υφίστανται αν το τεστ αυτό διαφοροποιεί (μέσω των ερωτημάτων-ερεθισμάτων του) μεταξύ πληθυσμών, *εκ των προτέρων*. Ένα τεστ επιλογής προσωπικού μπορεί να μεροληπτεί μεταξύ των δύο φύλων υπέρ των ανδρών, αν στα εξεταστικά αντικείμενα που χρησιμοποιεί για να καταλήξει σε αποτελέσματα για τους εξεταζόμενους περιλαμβάνονται δοκιμασίες που υποβοηθούνται (χωρίς αυτό να γίνεται αντιληπτό) από τη μυϊκή δύναμη ή το ύψος των ανδρών.

Γενικά, εάν για ένα οποιοδήποτε τεστ μπορούμε εκ των προτέρων να ορίσουμε κάποια ή κάποιες ομάδες ανθρώπων για τις οποίες το τεστ θα δίνει διαφορετικά αποτελέσματα τότε το τεστ αυτό είναι πολωμένο, ενώ το ίδιο πολωμένα είναι και τα αποτελέσματά του (συνήθεις παράμετροι πόλωσης είναι το φύλο, η φυλή, η χώρα και τα πολιτισμικά πλαίσια, οι κοινωνικές τάξεις, η ηλικία, οι γλωσσικές μειονότητες, κ.τ.λ.).

Ανίχνευση των πολώσεων ερωτημάτων

Είναι οι ευκολότερα εντοπιζόμενες πολώσεις μέσα σε ένα τεστ διότι μπορούμε να τις ανακαλύψουμε μέσω στατιστικών διαδικασιών. Ταυτόχρονα είναι και οι πιο ύπουλες καθώς δεν φαίνονται με την πρώτη ματιά και συνήθως δεν αποτελούν αντικείμενο ελέγχου των ερευνητών-δημιουργών του τεστ. Ένα παράδειγμα πόλωσης ερωτημάτων είναι αυτό που προέρχεται από μεταφράσεις και συνήθως οδηγεί σε προσαρμογές ξενόγλωσσων τεστ στην ελληνική, στην ινδική, στη γερμανική, κ.ο.κ. πραγματικότητα, δηλαδή το γεγονός ότι στα τεστ που έχουν δημιουργηθεί στις Η.Π.Α. περιλαμβάνονται συνήθως ερωτήσεις που αφορούν κυρίως στην παραγματικότητα των Η.Π.Α. (π.χ. πόσα *δολλάρια* κοστίζει κατά μέσον όρο ένα αυτοκίνητο). Στην ελληνική πραγματικότητα δεν θα πρέπει να περιμένουμε από ένα παιδί να γνωρίζει την αντίστοιχη απάντηση σε *δολλάρια*, αλλά θα προσαρμόσουμε την ερώτηση σε *δραχμές* (πλέον, σε *ευρώ*) . Στο Ελληνικό WISC-III, στην κλίμακα των Πληροφοριών το ερώτημα που αφορούσε στον *Graham Bell*

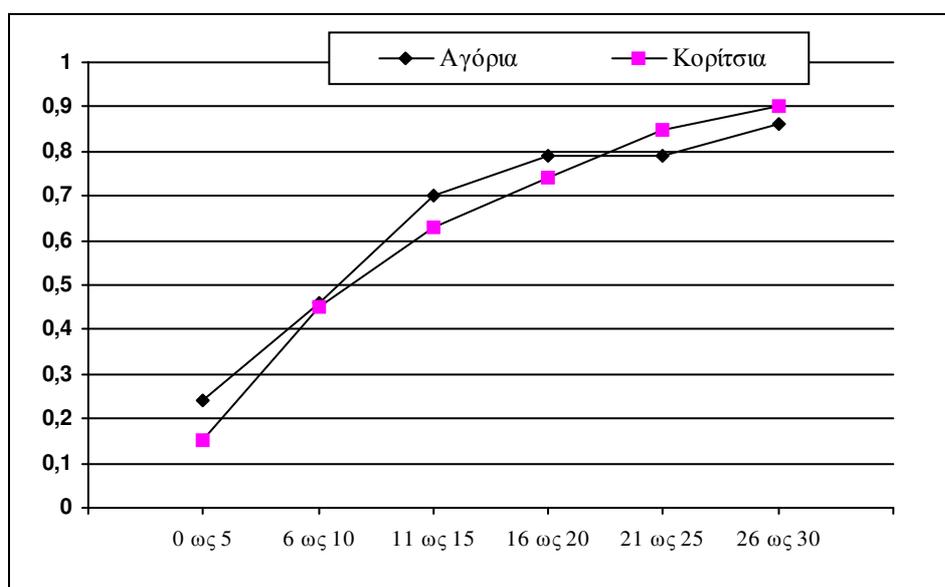
αντικατεστάθη από ερώτημα που αφορά στον *Γουτεμβέργιο* (που αναμένεται ότι είναι πιο γνώριμη μορφή στην ελληνική πραγματικότητα), η ερώτηση για την *πρωτεύουσα της Ελλάδας* (δύσκολη για έναν άγγλο, αλλά πολύ εύκολη για έναν έλληνα) με την ερώτηση για την *πρωτεύουσα της Πορτογαλίας* και η ερώτηση “*ποια είναι η απόσταση Λονδίνου-Εδιμβούργου*” με την ερώτηση “*ποια είναι η απόσταση Αθήνας-Θεσσαλονίκης*”.

Άλλη μορφή πόλωσης ερωτημάτων είναι οι διπλές αρνήσεις. Όταν ένα ερώτημα περιλαμβάνει διπλές αρνήσεις (π.χ. «Δεν τα καταφέρνω και πολύ άσχημα σε δύσκολες ερωτήσεις», με τη δεύτερη άρνηση ενσωματωμένη στο «*και πολύ άσχημα*») αποτελεί έτσι κι αλλιώς μεθοδολογικό σφάλμα. Η ερευνητική δραστηριότητα έχει επίσης δείξει για τις διπλές αρνήσεις ότι λειτουργούν και γίνονται αντιληπτές με διαφορετικό τρόπο σε διαφορετικούς πληθυσμούς (ανάλογα και με τη γραμματική δομή την οποία χρησιμοποιούν οι πληθυσμοί αυτοί ανεξαρτήτως γενικότερου ενδοπολιτιστικού πλαισίου). Σε έρευνα του *Scheuneman* το 1975, βρέθηκαν 7 πολωμένα ερωτήματα στο *Metropolitan Reading Test* για πληθυσμούς νέγων και λευκών, εκ των οποίων τα 6 περιείχαν διπλές αρνήσεις (π.χ. «*δείξε την εικόνα που δεν περιλαμβάνει ούτε μια γάτα ούτε ένα σκύλο*»¹⁴). Οι πολώσεις ήταν, σύμφωνα με τον ερευνητή, αποτέλεσμα της διαφορετικής γλωσσικής δομής και της χρήσης των διπλών αρνήσεων από τους δύο αυτούς πληθυσμούς.

Σε πολλές άλλες περιπτώσεις, βέβαια, η πόλωση στα ερωτήματα οφείλεται στο περιεχόμενο των ερωτημάτων αυτών. Συνήθως, τα ερωτήματα αυτά επηρεάζονται έντονα από παραμέτρους όπως η επιθυμία κοινωνικής αποδοχής ή δεν είναι έγκυρα ως προς την αντιπροσώπευση περιεχομένου (περιλαμβάνουν έννοιες γνωστές, προσφιλείς, συνήθεις, κ.ο.κ., μόνο σε τμήματα του πληθυσμού-στόχου). Ένα τέτοιο ερώτημα, μέσα σε ένα γενικότερο τεστ νοημοσύνης, θα ήταν ένα ερώτημα που, έστω για αγόρια και κορίτσια, θα είχε διαφορετικό δείκτη δυσκολίας *p*. Έστω ότι για τέσσερα τέτοια ερωτήματα οι δείκτες δυσκολίας για τα αγόρια είναι 0,86 , 0,74 , 0,70 και 0,62. Έστω, επίσης, ότι για τα ίδια ερωτήματα οι αντίστοιχοι δείκτες δυσκολίας για τα κορίτσια ίδιας ηλικίας είναι 0,85 , 0,79, 0,70 και 0,69. Όπως φαίνεται με την πρώτη ματιά, οι δείκτες δυσκολίας για το δεύτερο και το τέταρτο ερώτημα «*διαφοροποιούνται*» ανάλογα με το φύλο και πιθανώς τα ερωτήματα αυτά περιέχουν πολώσεις.

¹⁴ “*Mark the picture which shows neither a cat nor a dog.*”

Ένας σχετικά απλός τρόπος για την ανακάλυψη τέτοιων πολώσεων στα ύποπτα ερωτήματα είναι να επιχειρηθεί ανάλυση διακύμανσης με εξαρτημένη μεταβλητή την «επίδοση» σε κάθε ερώτημα. Ο ένας εκ των δύο ανεξάρτητων παραγόντων είναι οι ομάδες, έστω τα δύο φύλα. Η μεταβλητή που είναι πιθανόν να αλληλεπιδρά είναι τα διάφορα επίπεδα της επίδοσης στο σύνολο των ερωτημάτων ως να επρόκειτο για διαφορετικές ομάδες. Στο παρόν παράδειγμα τα επίπεδα αυτά θεωρητικά μπορεί να είναι 31 (εύρος συνολικής τιμής από 0 έως 30). Έστω ότι επιλέγουμε να διαφοροποιήσουμε ανά 5 τις επιδόσεις και να ορίσουμε 6 επίπεδα δυσκολίας. Στην περίπτωση αυτή, για κάποιο ερώτημα, έστω το 15ο, μπορεί να προκύψει η εικόνα που παρουσιάζεται στο ακόλουθο Σχήμα:



Υποθετικοί Μέσοι όροι επιτυχίας σε ένα ερώτημα ενός τεστ νοημοσύνης για αγόρια και κορίτσια ηλικίας 10 ετών, για κάθε επίπεδο συνολικής επίδοσης.

Στο πιο πάνω Σχήμα παρατηρούμε αυτό που οι *Van de Vijver & Leung* αποκαλούν μη-ομοιόμορφη πόλωση (*nonuniform bias*). Φαίνεται δηλαδή να μην υπάρχει ομοιόμορφη για τα δύο φύλα εικόνα «εξέλιξης» από το ένα επίπεδο συνολικής επίδοσης στα άλλο. Παράλληλα, υπάρχουν κάποιες διαφοροποιήσεις που «ευνοούν» τα αγόρια και κάποιες που «ευνοούν» τα κορίτσια. Ο αντίστοιχος συγκεντρωτικός υποθετικός Πίνακας της ανάλυσης διακύμανσης θα είχε ως εξής:

Πηγή διασποράς	df	SS	MS	F	p
Μεταξύ Ομάδων (Φύλο)	1	40	40	1,76	ΣΑ
Μεταξύ Ομάδων (Συνολική Επίδοση)	5	500	100	4,40	<0,01
Αλληλεπίδραση	5	400	80	3,52	<0,01
Εντός Ομάδων	88	2000	22,72		
Σύνολο	99	2940			

Από τον Πίνακα συνάγεται ότι η αλληλεπίδραση μεταξύ Φύλου και Επιπέδου Συνολικής Επίδοσης είναι στατιστικά σημαντική και το ίδιο συμβαίνει για την επίδραση των διαφορετικών επιπέδων συνολικής επίδοσης, ανεξαρτήτως Φύλου. Το τελευταίο είναι αναμενόμενο ως αποτέλεσμα (θα ήταν παράδοξο να μην υπάρχουν διαφορές) και δεν έχει ιδιαίτερη σημασία. Η στατιστικά σημαντική αλληλεπίδραση όμως δείχνει την ύπαρξη πολώσεων στο συγκεκριμένο ερώτημα. Το τελικό μας συμπέρασμα είναι ότι ο παράγοντας Φύλο πολώνει τις μετρήσεις μας και ή θα πρέπει να αναφερόμαστε ξεχωριστά στα δύο φύλα, ή θα πρέπει να αναθεωρήσουμε αυτό ή και άλλα αντίστοιχα πολωμένα ερωτήματα.

Η μέθοδος που περιγράψαμε έως εδώ έχει ένα σημαντικό μειονέκτημα: τον καθορισμό επιπέδων συνολικής επίδοσης με βάση κάποια σημεία τομής (0-5, 6-10, κ.ο.κ). Κάποιος άλλος μπορεί να καθόριζε τα σημεία αυτά με βάση τα δεκατημόρια (σύμφωνα με την παρατηρούμενη κατανομή) και κάποιος άλλος μπορεί να χρησιμοποιούσε δύο μόνο μεγάλες ομάδες (πάνω και κάτω από τη διάμεσο). Συνεπώς, η μέθοδος αυτή, αν και ιδιαίτερα χρήσιμη, εμπεριέχει το σοβαρό πρόβλημα της απώλειας πληροφοριών εξαιτίας του συγκεκριμένου συγκερασμού σε επίπεδα (με όποιον τρόπο και αν γίνει ο συγκερασμός αυτός). Εναλλακτικά, είναι δυνατόν να χρησιμοποιηθούν και άλλες, πιο λεπτομερείς και με μεγαλύτερες δυνατότητες πληροφόρησης μέθοδοι (μέθοδοι επαναληπτικών μετρήσεων, ανάλυση ψυχοδιαγραμμάτων, ανάλυση δομικής ισοτιμίας, κ.λπ.).

Αναλογιζόμενοι, για μία ακόμη φορά, την παράμετρο σφάλματος και τη θεωρία των αληθών τιμών, διαπιστώνουμε ότι στη συνολική διακύμανση των τιμών που παρατηρείται σε δειγματοληπτικές ψυχομετρικές διαδικασίες περιλαμβάνονται δύο ειδών λάθη: το σταθερό λάθος, που οφείλεται σε συγκεκριμένα χαρακτηριστικά του εκάστοτε δείγματος και, το διάσπαρτο λάθος, που οφείλεται σε τυχαίους παράγοντες και μόνο. Εάν κατορθώσουμε να βρούμε ότι η αληθής τιμή μαζί με το τυχαίο σφάλμα αποτελούν ιδιαίτερα μεγάλο τμήμα της συνολικής διακύμανσης, τότε, η ψυχομετρική διαδικασία που εφαρμόσαμε είναι αξιόπιστη. Εάν καταφέρουμε να δείξουμε ότι μόνο η αληθής τιμή, χωρίς την προσθήκη κανενός τμήματος λάθους ή θορύβου, αποτελεί μεγάλο τμήμα της συνολικής διακύμανσης των τιμών, τότε, το

τεστ είναι και έγκυρο¹⁵. Δυστυχώς, όμως, δεν μπορούμε να περιορίσουμε πολύ την τυχαία παράμετρο λάθους (μπορούμε να περιορίσουμε, ή τουλάχιστον να ελέγξουμε μόνον τη σταθερή παράμετρο λάθους) και έτσι, απόλυτη εγκυρότητα, όπως έχει ήδη συζητηθεί, δεν είναι δυνατό να υπάρξει.

¹⁵ Περισσότερες πληροφορίες μπορούν να αναζητηθούν στο σύγγραμμα: **I.N. Παρασκευόπουλος** (1993) *Μεθοδολογία Επιστημονικής Έρευνας*, Τόμος Β', Κεφάλαιο 6.

Μέρος 6. Έλεγχος ερωτημάτων

Στην προσπάθεια δημιουργίας και ελέγχου ενός οποιουδήποτε ψυχομετρικού εργαλείου, βασικό στάδιο αποτελεί ο έλεγχος των ερωτημάτων του εν λόγω εργαλείου, ώστε ο κατασκευαστής να είναι όσο το δυνατόν βέβαιος για την καταλληλότητα των ερωτημάτων που χρησιμοποιεί. Έχουμε ήδη συζητήσει τα θέματα της αξιοπιστίας και εγκυρότητας των ερωτημάτων, τα οποία είναι άμεσα συνδεδεμένα με τις διαδικασίες ελέγχου ερωτημάτων. Δεν μπορούμε να θεωρήσουμε τις διαδικασίες αυτές ως αυτόνομες ή άσχετες μεταξύ τους, καθώς η καθεμία είναι συμπληρωματική της άλλης. Τα ζητήματα της αξιοπιστίας και εγκυρότητας όμως, είναι γενικότερα και αποτελούν τη βάση για τον περαιτέρω έλεγχο των ερωτημάτων του ψυχομετρικού εργαλείου.

Θα πρέπει να διευκρινίσουμε ότι όταν μιλούμε για «ερώτημα», δεν αναφερόμαστε απαραίτητα σε μια ερωτηματική φράση για την οποία αναμένουμε κάποια απάντηση και η οποία μπορεί να αποτελεί τμήμα ενός γενικότερου συνόλου τέτοιων ερωτημάτων μέσα σε ένα ερωτηματολόγιο. Ο όρος αυτός από το σημείο αυτό και έπειτα θα χρησιμοποιείται με την έννοια του “item” δηλαδή του οποιουδήποτε ερευνητικού-ψυχομετρικού «ερεθίσματος» στο οποίο αναμένουμε αντίδραση από τον εξεταζόμενο. Είναι προφανές ότι η έννοια «ερέθισμα» περιλαμβάνει όλα τα είδη πιθανών μετρήσεων μέσω μιας ψυχομετρικής δοκιμασίας (ψυχοβιολογικές μετρήσεις, δοκιμασίες και μετρήσεις ικανοτήτων, νοημοσύνης, χαρακτηριστικών προσωπικότητας, στάσεων, αξιών, κ.τ.λ.). Επομένως τα αντίστοιχα ερεθίσματα-items μπορούν να έχουν δημιουργηθεί ανάλογα με το είδος της μέτρησης στο οποίο αποσκοπούν. Δεν θα ρωτήσουμε τον εξεταζόμενο αν το αυτόνομο νευρικό του σύστημα αυξάνει τους καρδιακούς παλμούς του κάτω από φοβικά ερεθίσματα, αλλά θα του παρουσιάσουμε κάποια φοβικά ερεθίσματα και θα μετρήσουμε την επίδρασή τους στην καρδιακή λειτουργία μέσω καρδιογράφου (χρειάζεται έγκριση της επιτροπής δεοντολογίας – *ethics committee*). Είναι όμως σημαντικό να θυμόμαστε ότι για κανένα σχεδόν ψυχομετρικό εργαλείο δεν μετρούμε την υπό μελέτη έννοια χρησιμοποιώντας μόνο ένα ερέθισμα. Η συνήθης τακτική είναι να χρησιμοποιήσουμε κλίμακα ερεθισμάτων-ερωτημάτων. Οι λόγοι για τη χρήση κλιμάκων είναι πολλοί, αλλά ένας από τους λόγους αυτούς είναι ότι για την αντίστοιχη κλίμακα ερεθισμάτων

επιτρέπεται (και επιβάλλεται) ο ψυχομετρικός έλεγχος μέσω των διαδικασιών ελέγχου ερωτημάτων.

Οι δύο συνήθεις μέθοδοι για τον έλεγχο ερωτημάτων είναι η Ανάλυση Ερωτημάτων και η Ανάλυση Παραγόντων. Και οι δύο αποσκοπούν στο να εξακριβώσουν τα πιθανά μειονεκτήματα κάποιων ερωτημάτων, να διαπιστώσουν τη σχέση μεταξύ των διαφόρων ερωτημάτων μιας κλίμακας και να καταλήξουν στα ερωτήματα που δείχνουν να ανταποκρίνονται θετικά στον έλεγχο αυτό, οδηγώντας έτσι στην τελική διαμόρφωση της ψυχομετρικής κλίμακας αξιολόγησης για το εργαλείο αυτό. Τις περισσότερες φορές, οι ψυχομέτρες δεν χρησιμοποιούν μόνο μία από τις δύο αυτές προσεγγίσεις αλλά συνδυάζουν τα αποτελέσματα που προκύπτουν από την καθεμιά, στην προσπάθειά τους να ανιχνεύσουν τα καλύτερα δυνατά ερωτήματα.

Ανάλυση Ερωτημάτων

Ο ψυχομέτρης έχει θέσει ως στόχο την κατασκευή μιας κλίμακας ερωτημάτων-ερεθισμάτων που περιλαμβάνει έναν συγκεκριμένο αριθμό εννοιών ενδιαφέροντος, οι οποίες έχουν προκύψει από τη βιβλιογραφία, από συνεντεύξεις, από προηγούμενες προσπάθειες δημιουργίας αντίστοιχων κλιμάκων αξιολόγησης και, φυσικά, από τα γενικότερα θεωρητικά πλαίσια. Είναι, επίσης, δυνατόν να υπάρχει έτοιμη κάποια αντίστοιχη ψυχομετρική κλίμακα αξιολόγησης, η οποία, όπως συμβαίνει πολύ συχνά στην ελληνική πραγματικότητα, δεν είναι «προσαρμοσμένη» στην πραγματικότητα αυτή, καθώς έχει κατασκευασθεί σε κάποια άλλη χώρα και αναφέρεται στον πληθυσμό της χώρας εκείνης και όχι της Ελλάδας. Ακόμη και στην περίπτωση αυτή, ο ψυχομέτρης που προσπαθεί να προσαρμόσει την υπάρχουσα κλίμακα στα δεδομένα της χώρας του, είναι αναγκασμένος να χρησιμοποιήσει την τεχνική της ανάλυσης ερωτημάτων.

Προϋποθέσεις

Η τεχνική αυτή στηρίζεται, μεταξύ άλλων, σε μερικές αναγκαίες προϋποθέσεις:

α) Κάθε ερώτημα της υπό χορήγηση ψυχομετρικής κλίμακας αξιολόγησης θα πρέπει να έχει μετρηθεί με την ίδια κλίμακα μέτρησης. Δεν έχει ιδιαίτερη σημασία

το κατά πόσο διαφορετικά ερωτήματα έχουν και διαφορετικό εύρος στις πιθανές απαντήσεις τους (για παράδειγμα, στην κλίμακα «Ομοιότητες» του Ελληνικού WISC-III υπάρχουν ερωτήματα στα οποία οι απαντήσεις βαθμολογούνται με 0 ή 1 και υπάρχουν και ερωτήματα για τα οποία οι απαντήσεις βαθμολογούνται με 0, 1 ή 2), αλλά έχει ιδιαίτερη σημασία οι μετρήσεις να γίνονται επί τη βάση της ίδιας μετρικής κλίμακας (για παράδειγμα, όλες οι μετρήσεις αυτές θα πρέπει να γίνουν με βάση την τακτική κλίμακα μέτρησης). Εάν, βέβαια, οι μετρήσεις μπορούν να αντιπροσωπεύσουν ένα αριθμητικό συνεχές και είναι δυνατό να χρησιμοποιηθεί η κλίμακα μέτρησης των ίσων διαστημάτων ή και ίσων λόγων, τότε αυτή η κλίμακα μέτρησης πρέπει να χρησιμοποιηθεί για την μέτρηση και καταγραφή όλων των ερωτημάτων της αντίστοιχης ψυχομετρικής κλίμακας-δοκιμασίας.

β) Η ψυχομετρική κλίμακα πρέπει να περιλαμβάνει αρκετά σε αριθμό ερωτήματα. Η προϋπόθεση αυτή ισχύει γενικότερα, αλλά κατά τη διαδικασία της ανάλυσης ερωτημάτων είναι απαραίτητο να έχουμε στη διάθεσή μας πληροφορίες για μεγάλο αριθμό ερωτημάτων ώστε να μπορούμε να επιλέξουμε τα καταλληλότερα. Εννοείται ότι εάν η αντιπροσωπευτικού περιεχομένου εγκυρότητα της δοκιμασίας απαιτεί, έστω, 50 ερωτήματα-έννοιες, τότε κατά τη διαδικασία ανάλυσης ερωτημάτων θα πρέπει να χρησιμοποιήσουμε τουλάχιστον 100 ερωτήματα, και αυτή είναι η συνήθης πρακτική κατά την προσαρμογή ενός ξενόγλωσσου τεστ στην εγχώρια πραγματικότητα.

γ) Οι κανόνες της τυχαίας δειγματοληψίας κατά τη συλλογή των δεδομένων είναι απαραίτητος όρος για την περαιτέρω ανάλυση. Εάν το δείγμα που έχουμε συλλέξει για την ανάλυση ερωτημάτων δεν είναι αντιπροσωπευτικό του στατιστικού πληθυσμού, τότε δεν μπορούμε να εμπιστευθούμε τα αποτελέσματα της ανάλυσης. Το ίδιο ισχύει και για κάθε ερευνητική προσπάθεια, ασχέτως της δημιουργίας κάποιου ψυχομετρικού μέσου, αλλά στην περίπτωση της ανάλυσης ερωτημάτων, η προϋπόθεση αυτή αποτελεί ακρογωνιαίο λίθο.

δ) Το δείγμα από το οποίο θα προκύψουν οι τιμές των ερωτημάτων που θα αναλυθούν είναι καλό να αποτελείται από μεγάλο, σχετικά, αριθμό ατόμων. Εάν για μια προκαταρκτική ερευνητική προσπάθεια χρειαζόμαστε έναν τουλάχιστον συμμετέχοντα περισσότερο από τον αριθμό των ερωτημάτων του μέσου συλλογής στοιχείων που χρησιμοποιούμε, στην περίπτωση της ανάλυσης ερωτημάτων καλό

είναι να χρησιμοποιήσουμε τουλάχιστον τέσσερις φορές περισσότερους συμμετέχοντες από τον αριθμό των ερωτημάτων της υπό ανάλυση κλίμακας αξιολόγησης. Πρέπει, βέβαια, να διευκρινισθεί ότι είναι δυνατόν ένα τεστ να περιλαμβάνει 10 αυθύπαρκτες τέτοιες κλίμακες ερωτημάτων με 40 ερωτήματα για κάθε κλίμακα αξιολόγησης. Αυτό δεν σημαίνει ότι το δείγμα θα πρέπει να είναι 1600 άτομα, αλλά 4 φορές περισσότερα από τον αριθμό ερωτημάτων της εκτενέστερης κλίμακας, στην περίπτωση 160 τουλάχιστον συμμετέχοντες. Η προϋπόθεση αυτή ισχύει για λόγους στατιστικής επεξεργασίας και θέτει ένα απόλυτο minimum. Το στατιστικό αυτό ζήτημα δεν αφορά τόσο στην ανάλυση ερωτημάτων αυτή καθ' εαυτή αλλά στον συνδυασμό των αποτελεσμάτων της με αυτά της ανάλυσης παραγόντων. Με τη λογική αυτή, αν κάποιος θέλει να παραμείνει στην ανάλυση ερωτημάτων και να μην προχωρήσει σε ανάλυση παραγόντων για τον περαιτέρω έλεγχο των ερωτημάτων της κλίμακας αξιολόγησης, δεν είναι αναγκαίος ο μεγάλος αριθμός συμμετεχόντων. Έτσι, πολλοί υποστηρίζουν ότι ένα δείγμα 100 συμμετεχόντων είναι αρκετό για την ανάλυση ερωτημάτων. Όμως από την άλλη μεριά, τίθεται και το ζήτημα της αντιπροσωπευτικότητας του δείγματος (σύμφωνα με την προϋπόθεση [γ]). Η ασφαλής λοιπόν προσέγγιση θα ήταν, ούτως ή άλλως, να συλλεγούν δεδομένα από ικανό αριθμό ατόμων. Ακόμη περισσότερο, πολλές φορές, λόγω της προϋπόθεσης της αντιπροσωπευτικότητας του δείγματος, δεν είναι αρκετή η κάλυψη μόνον των στατιστικών παραδοχών. Έτσι, στην περίπτωση του παραδείγματός μας, 160 συμμετέχοντες είναι μάλλον λίγοι και ο ψυχομέτρης δεν θα πρέπει να αρκεστεί στον αριθμό αυτόν, μόνο και μόνο επειδή καλύπτει την στατιστική προϋπόθεση, αλλά να συλλέξει στοιχεία από μεγαλύτερο δείγμα.

ε) Ο ψυχομέτρης θα πρέπει να τηρήσει όλες τις μεθοδολογικές και στατιστικές προϋποθέσεις, όπως αυτές προβλέπονται από τη θεωρία της Μεθοδολογίας Επιστημονικής Έρευνας¹⁶. Εφόσον η προσπάθεια ελέγχου ερωτημάτων και, στην περίπτωση, η προσπάθεια ανάλυσης ερωτημάτων γίνεται μέσω δειγματοληπτικής διαδικασίας, δεν θα πρέπει να παραβιάσουμε τις συνθήκες εκείνες που μας επιτρέπουν να καταλήξουμε σε συμπεράσματα για τα ερωτήματά μας.

¹⁶ Βλ. σύγγραμμα: **I.N. Παρασκευόπουλος** (1993) *Μεθοδολογία Επιστημονικής Έρευνας*, Τόμοι 1 και 2.

Στόχοι της ανάλυσης ερωτημάτων

Ο ψυχομέτρης, ξεκινώντας τη διαδικασία ανάλυσης ερωτημάτων, αναζητεί πληροφορίες που σχετίζονται με δύο, κυρίως, κριτήρια: τη συνάφεια κάθε μέτρησης με το συνολικό άθροισμα όλων των μετρήσεων, και τους δείκτες δυσκολίας των ερωτημάτων. Παράλληλα, ενδιαφέρεται και για την αξιοπιστία εσωτερικής συνέπειας που παρουσιάζει η κλίμακα αξιολόγησης, καθώς και για πιθανούς εξωγενείς παράγοντες που προκαλούν πολώσεις στα ερωτήματα αυτά.

Ο υπολογισμός του δείκτη δυσκολίας (*item facility*) για κάθε ερώτημα είναι το πρώτο βήμα. Ο δείκτης αυτός είναι το ποσοστό των συμμετεχόντων που απάντησαν στο ερώτημα αυτό ορθά. Ο αναγνώστης βέβαια αντιλαμβάνεται ότι για το συγκεκριμένο ζήτημα αναφερόμαστε σε «αντικειμενικά» τεστ και όχι σε τεστ αξιών, στάσεων, προσωπικότητας, κ.λπ. για τα οποία και δεν υπάρχουν «σωστές» ή «λάθος» απαντήσεις. Στα αντικειμενικά τεστ ο δείκτης δυσκολίας εκφράζει επιτυχία-αποτυχία και ταυτόχρονα τη διασπορά των τιμών των απαντήσεων¹⁷, ενώ στις υπόλοιπες περιπτώσεις τεστ, ο ίδιος δείκτης (καλείται *δείκτης προτίμησης*) εκφράζει την διασπορά των τιμών των απαντήσεων και μόνο. Επίσης, θα πρέπει να γνωρίζουμε ότι σε πολλά αντικειμενικά τεστ οι απαντήσεις δεν είναι δυωνυμικού τύπου (για παράδειγμα, οι απαντήσεις είναι δυνατό να βαθμολογούνται με βάση μια 10βαθμη κλίμακα μέτρησης). Στις περιπτώσεις αυτές, ο δείκτης δυσκολίας του κάθε ερωτήματος είναι ο αριθμητικός μέσος όρος για το ερώτημα αυτό και η διασπορά προκύπτει από τον στατιστικό δείκτη της τυπικής απόκλισης για το ερώτημα. Στις περιπτώσεις που οι απαντήσεις δίνονται με βάση πολλαπλές επιλογές, ισχύουν τα όσα ισχύουν και για τα αντικειμενικά τεστ, καθώς η τελική βαθμολόγηση κάθε απάντησης είναι δίτιμη.

Το συνολικό άθροισμα των επιμέρους τιμών για κάθε απάντηση του κάθε συμμετέχοντα αποτελεί τη συνολική ένδειξη για την υπό μελέτη έννοια. Αυτή η ένδειξη δεν είναι απαραίτητα και η πραγματική. Η διασπορά λάθους που εμπεριέχεται στα σύνολα αυτά προέρχεται από τη διασπορά λάθους για κάθε ερώτημα που με τη σειρά του προέρχεται από πολώσεις, τυχαίους και συστηματικούς εξωγενείς παράγοντες, κ.ο.κ. Θα πρέπει, λοιπόν, να έχουμε πάντα υπόψιν τη θεωρία των

¹⁷ Η διασπορά υπολογίζεται με βάση το δείκτη δυσκολίας ως εξής: $s^2 = p \otimes (1-p)$.

αληθών τιμών· εξάλλου, η όλη διαδικασία γίνεται για την εξακρίβωση της ύπαρξης ή όχι τέτοιων συστηματικών σφαλμάτων και την εξαίρεση των ερωτημάτων εκείνων που επηρεάζονται από τα σφάλματα αυτά. Σε πρώτη όμως φάση, το σύνολο των τιμών μπορεί να συνδυαστεί με τις απαντήσεις στις επιμέρους ερωτήσεις που συντελούν ή όχι στη διαμόρφωση του συνόλου αυτού. Ο στόχος είναι η συσχέτιση των επιμέρους απαντήσεων στα ερωτήματα με το συνολικό, ενδεικτικό άθροισμά τους. Η εκτίμηση των συσχετίσεων θα οδηγήσει σε συμπεράσματα σχετικά με την χρησιμότητα του κάθε ερωτήματος. Εάν, για παράδειγμα μια τέτοια συσχέτιση είναι της τάξης του 0,65, τότε το αντίστοιχο ερώτημα φαίνεται να έχει αρκετά μεγάλη σχέση με το τελικό άθροισμα, εάν όμως ο αντίστοιχος δείκτης είναι 0,23, τότε η σχέση μεταξύ ερωτήματος και τελικού αθροίσματος δεν είναι ικανοποιητική και θα πρέπει ίσως το ερώτημα αυτό να αντικατασταθεί.

Παράλληλα, η αναζήτηση της αξιοπιστίας εσωτερικής συνέπειας της κλίμακας ερωτημάτων έχει στόχο την εξαίρεση ερωτημάτων που προκαλούν μεγάλη διασπορά λάθους στο τελικό σύνολο των τιμών και τα οποία δεν παρουσιάζουν υψηλή συνάφεια με τα υπόλοιπα ερωτήματα της κλίμακας. Ο στατιστικός στόχος εδώ είναι να μεγιστοποιηθεί ο δείκτης *Cronbach α* (ή ο δείκτης KR-20 εάν πρόκειται για απαντήσεις δυωνυμικού τύπου)· αντίστοιχη προσέγγιση μπορεί κανείς να εφαρμόσει και κατά τη διαδικασία της ανάλυσης παραγόντων, όπως θα συζητηθεί σε επόμενο κεφάλαιο. Ταυτόχρονα, ο ψυχομέτρης μπορεί να ελέγξει το κατά πόσο υπάρχουν πολώσεις στις απαντήσεις των διαφόρων ερωτημάτων της κλίμακας, καθώς έχει τη δυνατότητα να ελέγξει την πιθανότητα στατιστικής αλληλεπίδρασης μεταξύ των χαρακτηριστικών των συμμετεχόντων και των ερωτημάτων, αλληλεπίδραση που προκύπτει όταν το δείγμα αποτελείται από δύο ή περισσότερους ανομοιογενείς πληθυσμούς ή όταν υπάρχει συστηματικός λόγος για τον οποίο οι συμμετέχοντες απαντούν με τον Α ή Β τρόπο (για παράδειγμα, ένα δείγμα φοιτητών του Οικονομικού θα απαντούσε με μεγάλη ευκολία σε ερωτήματα μαθηματικού τύπου και με μεγάλη δυσκολία σε ερωτήματα φιλολογικού χαρακτήρα). Στην περίπτωση τέτοιων πολώσεων, ο ψυχομέτρης δεν έχει τη στατιστική δυνατότητα αθροίσματος των επιμέρους απαντήσεων σε μία συνολική τιμή προτού προβεί σε μαθηματικές (συνήθως μη-γραμμικές) μετατροπές των τιμών που έχει συλλέξει. Όπως και να είναι

όμως τα πράγματα, ο ψυχομέτρης είναι ενήμερος για τις δυνατότητες των τιμών που έχει στα χέρια του.

Διαδικασία ανάλυσης ερωτημάτων

Έχοντας τα παραπάνω ως βασικούς στόχους, προχωρούμε στα ακόλουθα βήματα:

- 1) Υπολογίζουμε για κάθε συμμετέχοντα το συνολικό άθροισμα των επιμέρους απαντήσεών του για όλα τα ερωτήματα της ψυχομετρικής κλίμακας αξιολόγησης.
- 2) Υπολογίζουμε, για κάθε ερώτημα τον δείκτη δυσκολίας του ερωτήματος, ο οποίος εκφράζεται είτε ως σχετική συχνότητα, είτε ως αριθμητικός μέσος όρος και διατάσσουμε-ιεραρχούμε τα ερωτήματα σε αύξουσα σειρά με βάση τον δείκτη αυτόν.
- 3) Υπολογίζουμε όλες τις συνάφειες μεταξύ των επιμέρους απαντήσεων και του συνολικού τους αθροίσματος (ανάλογα με την κλίμακα μέτρησης, υπολογίζονται και οι κατάλληλοι δείκτες συνάφειας).
- 4) Υπολογίζουμε τον δείκτη αξιοπιστίας εσωτερικής συνέπειας και τους αντίστοιχους δείκτες που θα προέκυπταν αν αφαιρούσαμε κάθε φορά και από ένα ερώτημα της κλίμακας. Ταυτόχρονα, υπολογίζουμε τον δείκτη δυνατότητας άθροισης των επιμέρους τιμών των απαντήσεων των συμμετεχόντων (*Tukey's test for non-additivity*) και εάν θεωρηθεί αναγκαίο, προχωρούμε σε μαθηματικές μετατροπές των τιμών, επανυπολογίζουμε το άθροισμα και επαναλαμβάνουμε τη διαδικασία εξ αρχής.

Για να γίνει περισσότερο κατανοητή η διαδικασία αυτή, θα χρησιμοποιηθεί ένα παράδειγμα που αφορά στις απαντήσεις 956 παιδιών ηλικίας 6 έως 16 ετών για την ψυχομετρική κλίμακα «Ελληνικό WISC-III» και την υποκλίμακά του «Συμπλήρωση Εικόνων».

Στατιστικοί δείκτες για την κλίμακα «Συμπλήρωση Εικόνων»
του Ελληνικού WISC-III, όπως προέκυψαν από το δείγμα της στάθμησης (N=956)

Μέτρηση	Μέσος όρος	Τυπικό σφάλμα μέτρησης Μ.Ο.	Τυπική απόκλιση	Διασπορά	Minimum	Maximum
ΣΕ1	0,99	≈0,00	0,07	0,01	0,00	1,00
ΣΕ2	0,99	≈0,00	0,10	0,01	0,00	1,00
ΣΕ3	0,99	≈0,00	0,12	0,01	0,00	1,00
ΣΕ4	0,95	≈0,01	0,22	0,05	0,00	1,00
ΣΕ5	0,96	≈0,01	0,20	0,04	0,00	1,00
ΣΕ6	0,95	≈0,01	0,22	0,05	0,00	1,00
ΣΕ7	0,88	≈0,01	0,32	0,11	0,00	1,00
ΣΕ8	0,93	≈0,01	0,25	0,06	0,00	1,00
ΣΕ9	0,79	≈0,01	0,40	0,16	0,00	1,00
ΣΕ10	0,78	≈0,01	0,41	0,17	0,00	1,00
ΣΕ11	0,80	≈0,01	0,40	0,16	0,00	1,00
ΣΕ12	0,78	≈0,01	0,42	0,17	0,00	1,00
ΣΕ13	0,81	≈0,01	0,39	0,15	0,00	1,00
ΣΕ14	0,77	≈0,01	0,42	0,18	0,00	1,00
ΣΕ15	0,65	≈0,02	0,48	0,23	0,00	1,00
ΣΕ16	0,73	≈0,01	0,44	0,20	0,00	1,00
ΣΕ17	0,57	≈0,02	0,50	0,25	0,00	1,00
ΣΕ18	0,64	≈0,02	0,48	0,23	0,00	1,00
ΣΕ19	0,53	≈0,02	0,50	0,25	0,00	1,00
ΣΕ20	0,44	≈0,02	0,50	0,25	0,00	1,00
ΣΕ21	0,41	≈0,02	0,49	0,24	0,00	1,00
ΣΕ22	0,42	≈0,02	0,49	0,24	0,00	1,00
ΣΕ23	0,27	≈0,01	0,45	0,20	0,00	1,00
ΣΕ24	0,24	≈0,01	0,43	0,18	0,00	1,00
ΣΕ25	0,21	≈0,01	0,41	0,17	0,00	1,00
ΣΕ26	0,22	≈0,01	0,41	0,17	0,00	1,00
ΣΕ27	0,14	≈0,01	0,35	0,12	0,00	1,00
ΣΕ28	0,11	≈0,01	0,31	0,10	0,00	1,00
ΣΕ29	0,10	≈0,01	0,30	0,09	0,00	1,00
ΣΕ30	0,05	≈0,01	0,22	0,05	0,00	1,00

Παρατηρούμε στον πιο πάνω Πίνακα ότι οι απαντήσεις των συμμετεχόντων για κάθε ερώτημα έχουν βαθμολογηθεί με 0 ή 1 (είναι δυωνυμικού τύπου). Κατά συνέπεια οι αριθμητικοί μέσοι όροι εκφράζουν κατ' ουσίαν ποσοστά επιτυχίας (και ποσοστά αποτυχίας (δείκτες p) αν αφαιρέσουμε τους αριθμούς αυτούς από τη μονάδα). Έτσι, παρατηρούμε ότι με κάποιες μικρές εξαιρέσεις, η σειρά χορήγησης των ερωτημάτων (1 έως 30) ταυτίζεται με την αύξουσα σειρά σύμφωνα με τον δείκτη δυσκολίας. Αυτό συμβαίνει διότι για τη στάθμηση του Ελληνικού WISC-III είχαν προηγηθεί προκαταρκτικές έρευνες για τις οποίες και έγιναν αναλύσεις ερωτημάτων, από τις οποίες προέκυψε και η τελική σειρά χορήγησης που δόθηκε στο δείγμα της στάθμησης. Με άλλα λόγια, η πιο πάνω εικόνα είναι σχεδόν η ιδανική, τουλάχιστον ως προς τους δείκτες δυσκολίας και σπάνια προκύπτει με την πρώτη προσπάθεια.

Ένα άλλο στοιχείο που παρατηρούμε στον Πίνακα είναι ότι για πολύ μικρά ή πολύ μεγάλα ποσοστά αποτυχίας η διασπορά των τιμών είναι πολύ μικρή. Αυτό συμβαίνει διότι οι κατανομές συχνότητας για τις αντίστοιχες μετρήσεις παρουσιάζουν οξείες ασυμμετρίες (δεξιά ή αριστερά) και είναι, ταυτόχρονα, λεπτόκυρτες. Οι μόνες περιπτώσεις που οι κατανομές συχνότητας δεν είναι ασύμμετρες και λεπτόκυρτες είναι εκείνες στις οποίες τα παιδιά έχουν απαντήσει ορθά στην ερώτηση περίπου κατά το ήμισυ. Στις περιπτώσεις αυτές οι κατανομές συχνότητας είναι τουλάχιστον ισουψείς (δεν μπορούν να είναι συμμετρικές κωδωνοειδείς, καθώς έχουμε απαντήσεις δυωνυμικού τύπου) και παρουσιάζουν τη μέγιστη δυνατή στην περίπτωση διασπορά τιμών (περίπου 0,25). Αντίστοιχα, βέβαια, διαμορφώνονται και η τυπική απόκλιση καθώς και το τυπικό σφάλμα μέτρησης του μέσου όρου για κάθε ερώτημα.

Από τον Πίνακα αυτόν προκύπτουν δύο συμπεράσματα: α) ότι η σειρά χορήγησης των ερωτημάτων, με κάποιες μικρές αναπροσαρμογές, αντιστοιχεί σε σταδιακή αύξηση της δυσκολίας, κάτι που ικανοποιεί το πρώτο κριτήριο για την ανάλυση ερωτημάτων¹⁸ και, β) ότι 21 από τα 30 ερωτήματα (70%) παρουσιάζουν διασπορά μεγαλύτερη ή ίση του 0,10 και 9 ερωτήματα από τα 30 (30%) παρουσιάζουν διασπορά μεγαλύτερη ή ίση του 0,20. Τα ερωτήματα αυτά διαθέτουν ικανή διαφοροποιητική ισχύ, δηλαδή μπορούν να προσθέσουν λεπτομερείς πληροφορίες για την συνολική διαφοροποίηση ως προς τη νοημοσύνη του κάθε εξεταζόμενου. Η έννοια της διαφοροποιητικής ισχύος θα αναλυθεί περαιτέρω σε επόμενα κεφάλαια.

Το επόμενο στάδιο είναι ο υπολογισμός των συναφειών των τιμών απάντησης σε κάθε ερώτημα με τον συνολικό αθροιστικό δείκτη. Στην περίπτωση, ο δείκτης αυτός θα είναι δείκτης *point-biserial*, καθώς η μία από τις δύο μετρήσεις είναι δίτιμη και η άλλη είναι αριθμητική μέτρηση. Ο υπολογισμός των δεικτών αυτών καταλήγει στον ακόλουθο πίνακα:

¹⁸ Οι λόγοι της αναγκαιότητας αυτής θα αναλυθούν στο κεφάλαιο της θεωρίας των λανθανόντων χαρακτηριστικών.

Δείκτες συνάφειας ερωτημάτων-συνόλου

Σύνολο τιμών με: <i>point-biserial</i>	ΣΕ1 0,1636	ΣΕ2 0,1999	ΣΕ3 0,2227	ΣΕ4 0,4186	ΣΕ5 0,2837	ΣΕ6 0,3370
Σύνολο τιμών με: <i>point-biserial</i>	ΣΕ7 0,3545	ΣΕ8 0,3599	ΣΕ9 0,4910	ΣΕ10 0,5563	ΣΕ11 0,5108	ΣΕ12 0,5311
Σύνολο τιμών με: <i>point-biserial</i>	ΣΕ13 0,5113	ΣΕ14 0,5855	ΣΕ15 0,5837	ΣΕ16 0,6194	ΣΕ17 0,5968	ΣΕ18 0,6329
Σύνολο τιμών με: <i>point-biserial</i>	ΣΕ19 0,6325	ΣΕ20 0,5817	ΣΕ21 0,5764	ΣΕ22 0,6047	ΣΕ23 0,4883	ΣΕ24 0,5138
Σύνολο τιμών με: <i>point-biserial</i>	ΣΕ25 0,4492	ΣΕ26 0,4969	ΣΕ27 0,3923	ΣΕ28 0,4058	ΣΕ29 0,3898	ΣΕ30 0,2807

Από τον πιο πάνω Πίνακα προκύπτει ότι σχεδόν όλες οι επιμέρους μετρήσεις (απαντήσεις στα 30 ερωτήματα της κλίμακας «Συμπλήρωση Εικόνων») έχουν μέτρια ως υψηλή συνάφεια με το συνολικό άθροισμά τους. Εξαιρώντας τις πρώτες τρεις επιμέρους μετρήσεις, οι οποίες και λόγω μηδενικής διασποράς καταλήγουν σε πολύ χαμηλούς δείκτες συνάφειας, καταλήγουμε στο ότι όλα τα επιμέρους ερωτήματα προσφέρουν, κατά τα προσδοκώμενα, στο συνολικό άθροισμα. Δεν παρατηρείται καμία αρνητική συνάφεια, η οποία και θα σήμαινε ότι τα αντίστοιχα ερωτήματα μετρά αντίστροφες έννοιες και είναι πιθανώς ανέγκυρο, ενώ δεν παρατηρούνται επίσης πολύ υψηλές συνάφειες, οι οποίες και θα σήμαιναν ότι το συνολικό άθροισμα θα μπορούσε να προκύψει μόνο από αυτές τις επιμέρους ερωτήσεις και επομένως οι υπόλοιπες θα ήταν αδρανείς έως ανούσιες.

Σύμφωνα με τα παραπάνω, τα δύο βασικά κριτήρια της ανάλυσης ερωτημάτων πληρούνται και απομένει ο έλεγχος της αξιοπιστίας εσωτερικής συνέπειας της κλίμακας αυτής μαζί με τον έλεγχο της δυνατότητας άθροισης των επιμέρους μετρήσεων και της χρήσης του συνολικού αυτού αθροίσματος ως γενικού δείκτη για την κλίμακα «Συμπλήρωση Εικόνων».

Σύμφωνα με τον Πίνακα της σελ. 68 καθώς και την ανάλυση διακύμανσης και τον δείκτη αξιοπιστίας στον οποίο καταλήγουμε είναι δυνατόν να προβούμε στα ακόλουθα συμπεράσματα για την κλίμακα 'Συμπλήρωση Εικόνων': α) Η αξιοπιστία εσωτερικής συνέπειας που παρουσιάζει η κλίμακα αυτή είναι ιδιαίτερα υψηλή, ενώ κανένα ερώτημα δεν φαίνεται να εισάγει ιδιαίτερα μεγάλη διασπορά θορύβου καθώς

Αποτελέσματα ανάλυσης αξιοπιστίας

Ερώτημα	Μ.Ο. κλίμακας αν το ερώτημα διαγραφεί	Διασπορά κλίμακας αν το ερώτ. διαγραφεί	‘Διορθωμένη’ συνάφεια ερωτημάτων-συνόλου	Δείκτης α αν το ερώτημα διαγραφεί
ΣΕ1	17,1318	29,8088	0,1506	0,8872
ΣΕ2	17,1370	29,6891	0,2103	0,8868
ΣΕ3	17,1412	29,5957	0,2466	0,8864
ΣΕ4	17,1757	28,9262	0,4124	0,8841
ΣΕ5	17,1663	29,3032	0,2794	0,8858
ΣΕ6	17,1757	29,1063	0,3340	0,8851
ΣΕ7	17,2458	28,7573	0,3077	0,8853
ΣΕ8	17,1925	28,9744	0,3355	0,8849
ΣΕ9	17,3316	27,9245	0,4322	0,8830
ΣΕ10	17,3421	27,5468	0,5133	0,8811
ΣΕ11	17,3264	27,7991	0,4678	0,8822
ΣΕ12	17,3494	27,6265	0,4873	0,8817
ΣΕ13	17,3117	27,8567	0,4693	0,8822
ΣΕ14	17,3556	27,3708	0,5421	0,8804
ΣΕ15	17,4749	27,0915	0,5268	0,8808
ΣΕ16	17,3923	27,0931	0,5748	0,8795
ΣΕ17	17,5586	26,9212	0,5378	0,8805
ΣΕ18	17,4843	26,8029	0,5839	0,8792
ΣΕ19	17,6004	26,6904	0,5798	0,8793
ΣΕ20	17,6851	27,0076	0,5187	0,8810
ΣΕ21	17,7207	27,0350	0,5199	0,8810
ΣΕ22	17,7061	26,9114	0,5420	0,8804
ΣΕ23	17,8536	27,7335	0,4261	0,8833
ΣΕ24	17,8828	27,6596	0,4623	0,8823
ΣΕ25	17,9184	28,0897	0,3897	0,8840
ΣΕ26	17,9079	27,8365	0,4412	0,8828
ΣΕ27	17,9843	28,5327	0,3422	0,8848
ΣΕ28	18,0188	28,6572	0,3550	0,8845
ΣΕ29	18,0241	28,7439	0,3370	0,8848
ΣΕ30	18,0753	29,3032	0,2431	0,8861

Ανάλυση Διακύμανσης

Πηγή	Άθροισμα Τετραγ.	Βαθμοί Ελευθερίας	Μέσα Τετραγ.	F	p
Μεταξύ συμμετεχόντων	952,8562	955	0,9978		
Εντός των μετρήσεων	5905,6333	27724	0,2130		
Μεταξύ μετρήσεων	2770,6872	29	95,5409	844,0356	0,0000
Θόρυβος	3134,9461	27695	0,1132		
Μη-Άθροισμότητα	19,4071	1	19,4071	172,5093	0,0000
Ισορροπία	3115,5391	27694	0,1125		
Σύνολο	6858,4895	28679	0,2391		

Γενικός μέσος όρος ερωτημάτων: 0,6042

Εκτίμηση Tukey για τη δύναμη στην οποία πρέπει να υψωθούν οι επιμέρους τιμές, ώστε να επιτευχθεί αθροισμότητα = 1,2774

Hotelling's $T^2 = 25740,7010$ F = 861,5862 Prob. = 0,0000
 Βαθμοί ελευθερίας: Αριθμητής= 29 Παρονομαστής = 927

Δείκτης αξιοπιστίας για 30 ερωτήματα
 $\alpha = 0,8865$

α διορθωμένο για τον πληθυσμό = 0,8809

το α δεν αλλάζει αν διαγραφούν ερωτήματα. β) Ο μέσος όρος της κλίμακας στο σύνολό της δεν μεταβάλλεται ιδιαίτερα αν διαγραφούν ερωτήματα, ενώ το ίδιο συμβαίνει και για τη διασπορά της κλίμακας. Αυτή η διαπίστωση γενικεύεται στον συνολικό δείκτη που προκύπτει από την κλίμακα, ο οποίος και αντιπροσωπεύεται από τον μέσο όρο καθώς και τη διασπορά της κλίμακας. γ) Οι διορθωμένοι δείκτες συνάφειας ερωτημάτων-συνόλου δεν μεταβάλλονται ιδιαίτερα όταν κάποιο ερώτημα διαγραφεί. δ) Η στατιστικά σημαντική διαφορά που παρατηρείται μεταξύ των μετρήσεων αντιπροσωπεύει τις γενικές διαφορές στους δείκτες δυσκολίας, επομένως η κλίμακα φαίνεται να αποτελείται από ερωτήματα διαφορετικής δυσκολίας. Το συμπέρασμα αυτό επιβεβαιώνεται και από το πολυμεταβλητό κριτήριο Hotelling's T^2 το οποίο παράλληλα με τις πολυμεταβλητές διαφορές μεταξύ των ερωτημάτων δείχνει και ότι οι διαφορές αυτές είναι «διαφορετικές» και μεταξύ των συμμετεχόντων (τα ψυχοδιαγράμματα για τα 30 αυτά ερωτήματα δεν είναι παράλληλα). ε) Ο δείκτης μη-αθροισιμότητας είναι στατιστικά σημαντικός και δείχνει ότι υπάρχει αλληλεπίδραση συμμετεχόντων-ερωτημάτων, η οποία, στην περίπτωση, προφανώς προέρχεται από το γεγονός ότι στα 956 αυτά άτομα μετέχουν ηλικίες 6 έως 16 ετών, επομένως είναι πιθανόν (και αναμενόμενο εξάλλου) κάποια ερωτήματα να είναι πιο εύκολα για άλλους και για άλλους, πιο δύσκολα. Το επόμενο βήμα θα ήταν να επαναληφθεί η ανάλυση αξιοπιστίας για κάθε ηλικία ξεχωριστά και να επανεκτιμηθεί ο δείκτης Tukey. Στη φάση αυτή, καθώς η αλγεβρική δύναμη που προτείνει ο Tukey για την επίτευξη αθροισιμότητας είναι πολύ κοντά στη μονάδα, και καθώς η αιτία είναι η αναμενόμενες διαφορές μεταξύ των ηλικιών, δεν έγινε μαθηματική μετατροπή. Εάν, όμως, η δύναμη αυτή ήταν -1 τότε θα υπολογίζαμε τον αντίστροφο

($1 \div X$) κάθε αρχικής τιμής και θα επανυπολογίζαμε το συνολικό άθροισμα, κ.ο.κ.

Αποτελέσματα και περαιτέρω διαδικασία

Ανακεφαλαιωτικά, για την υπό συζήτηση κλίμακα του Ελληνικού WISC-III «Συμπλήρωση Εικόνων», η ανάλυση ερωτημάτων έδειξε ότι τα ερωτήματα έχουν τοποθετηθεί σε σειρά δυσκολίας, έχουν διαφορετικούς μεταξύ τους δείκτες δυσκολίας και οι συνάφειες με το τελικό σύνολο είναι ικανοποιητικές. Τέλος, η αξιοπιστία εσωτερικής συνέπειας είναι επίσης ικανοποιητική, αν και θα πρέπει να λάβουμε

υπόψιν ότι υφίστανται αλληλεπιδράσεις συμμετεχόντων και ερωτημάτων αλλά και ότι η επίδραση αυτή είναι και αναμενόμενη αλλά και σχετικά ακίνδυνη στην περίπτωση. Κατά συνέπεια, η κλίμακα αυτή αξιολόγησης αποδεικνύεται έτοιμη προς γενικότερη χρήση. Σε αντίθετη περίπτωση θα αφαιρούσαμε τα «προβληματικά» ερωτήματα, θα επαναλαμβάνουμε τις διαδικασίες της ανάλυσης ερωτημάτων και ίσως επανασυλλέγαμε στοιχεία από νέο δείγμα για να επαναλάβουμε, για μια ακόμη φορά, την ανάλυση ερωτημάτων.

Σύμφωνα με τον γενικό κανόνα, είναι συνήθως απαραίτητη η επανάληψη της διαδικασίας, όταν πλέον έχουμε καταλήξει στα ερωτήματα που φαίνεται να πληρούν τα κριτήρια, για κάθε φύλο ξεχωριστά. Οι διαφορές φύλου πολλές φορές αποτελούν την παράμετρο που προκαλεί την αλληλεπίδραση συμμετεχόντων και ερωτημάτων και γενικότερα εισάγουν ανεξέλεγκτη διασπορά η οποία και δεν θα επηρεάζει τα αποτελέσματα όταν η ανάλυση ερωτημάτων γίνει για κάθε φύλο ξεχωριστά. Εάν προκύψουν διαφορετικά ερωτήματα για κάθε φύλο τότε πρέπει να βρεθούν τα κοινά, να διπλασιαστούν σε αριθμό και να επαναληφθούν όλες οι διαδικασίες από την αρχή.

Στην ανάλυση ερωτημάτων πρέπει να προσεχθούν δύο ακόμη σημεία. Το πρώτο είναι ότι υπάρχει η περίπτωση ο τελικός αριθμός των ερωτημάτων που φαίνεται να «λειτουργούν» ικανοποιητικά να είναι χαμηλός. Στην περίπτωση αυτή πρέπει να δημιουργήσουμε νέο κατάλογο ερωτημάτων και να επαναλάβουμε τη διαδικασία. Το δεύτερο ζήτημα αφορά στην πιθανή ύπαρξη, μέσα στο σύνολο των χρησιμοποιούμενων ερωτημάτων, υποσυνόλων που αναφέρονται σε διαφορετικές μεταξύ τους έννοιες. Είναι δυνατόν, παρόλο που η ανάλυση ερωτημάτων και η ανάλυση αξιοπιστίας δείχνουν να καταλήγουν σε ένα σύνολο ερωτημάτων που μετρά και αναφέρεται σε *μία* έννοια, να ενυπάρχουν περισσότερες από μία έννοιες στο σύνολο αυτό. Στην περίπτωση αυτή είναι πιθανό να νομίζουμε ότι προσεγγίζουμε μία μέτρηση (αυτή που μας ενδιαφέρει εξ αρχής) και τελικά να προσεγγίζουμε ταυτόχρονα δύο ή και περισσότερες έννοιες. Ακόμη και όταν δεν υπάρχουν υπόνοιες για κάτι τέτοιο, είναι χρήσιμο ως απαραίτητο να εξετάσουμε τα συλλεγένητα στοιχεία και μέσω ανάλυσης παραγόντων. Αυτή είναι και η συνήθης πρακτική και εφόσον προκύψουν περισσότερες από μία «διαστάσεις» μέσα στο σύνολο των ερωτημάτων, τότε επαναλαμβάνουμε την ανάλυση παραγόντων ενσωματώνοντας και άλλα στοιχεία που προκύπτουν από άλλη ή άλλες σχετικές κλίμακες και προσπαθούμε να

διακρίνουμε τη δική μας κλίμακα μέσα στο γενικότερο σύνολο. Εάν αυτό συμβεί, τότε δεν υπάρχει πρόβλημα. Αλλιώς, η κλίμακά μας, όσο και αν δείχνει αυτόνομη, δεν είναι, αλλά σχετίζεται με άλλες διαστάσεις, ενώ ταυτόχρονα περιλαμβάνει και μετρά περισσότερες της μιας έννοιες. Αναμφίβολα, η τελευταία περίπτωση δεν είναι καταστροφική και μάλιστα είναι πολύ συνηθισμένη. Η ιδανική όμως περίπτωση είναι μια κλίμακα αξιολόγησης να αναφέρεται σε μία μόνο έννοια-«παράγοντα». Η ανάλυση ερωτημάτων, εν τέλει, πολλές φορές καταλήγει στα ίδια αποτελέσματα με αυτά της ανάλυσης παραγόντων και δεν είναι απαραίτητο ότι δεν θα διακρίνουμε και από την ανάλυση ερωτημάτων την ύπαρξη περισσότερων της μιας διαστάσεων. Από την άλλη, η ανάλυση ερωτημάτων διαθέτει το πλεονέκτημα ότι ξεκαθαρίζει στην αρχή της διαδικασίας τα «ύποπτα» ερωτήματα τα οποία και αποβάλλονται πριν την επιβεβαιωτική επανάληψη της διαδικασίας και την παράλληλη χρήση της ανάλυσης παραγόντων.

Τέλος, υπάρχουν διάφορες άλλες μέθοδοι ανάλυσης ερωτημάτων. Ανωτέρω περιγράψαμε μόνο το γενικό πλαίσιο. Ανάλογα με τις ιδιαίτερες συνθήκες για κάθε διαδικασία ανάλυσης ερωτημάτων, εφαρμόζονται και εναλλακτικές αναλυτικές στρατηγικές. Η πιο συνήθης είναι ο υπολογισμός των δεικτών διαφοροποίησης, μια μέθοδος που καταλήγει σε έναν δείκτη για κάθε ερώτημα-δοκιμασία που επιβεβαιώνει ή όχι τη διαφοροποιητική δύναμη του item αυτού. Ο υπολογισμός των δεικτών αυτών γίνεται με διάφορες μεθόδους, όπως η διάκριση του συνολικού δείγματος σε τρεις ομάδες με βάση τη συνολική τιμή-άθροισμα (το υψηλότερο 27%, το χαμηλότερο 27%, και το ενδιάμεσο 46%). Οι δύο ακραίες ομάδες συγκρίνονται με βάση τους δείκτες δυσκολίας που τους αντιστοιχούν σε κάθε item και για το item αυτό προκύπτει ο δείκτης διαφοροποίησης. Άλλη μέθοδος είναι η σύγκριση των μέσων όρων των τριών ομάδων μέσω ανάλυσης διακύμανσης (παραμετρικής ή μη-παραμετρικής) ή και πιο πολύπλοκες μέθοδοι όπως αυτή της “λογιστικής” παλινδρόμησης (*logistic regression*) ή της ανάλυσης διαφοροποιητικής ισχύος (*discriminant function analysis*).

Μία άλλη προσέγγιση στην ανάλυση ερωτημάτων γίνεται μέσω της ιεραρχικής ανάλυσης συστάδων (*hierarchical cluster analysis*), όπου ο στόχος είναι να ανιχνευθούν ομοιογενή σύνολα ερωτημάτων. Η ομοιογένειά τους αυτή θα καθορίσει επιμέρους υποομάδες ερωτημάτων μέσα στο σύνολο της ψυχομετρικής

κλίμακας αξιολόγησης και οι οποίες θα μπορούν να ελεγχθούν περαιτέρω για τα ψυχομετρικά τους χαρακτηριστικά. Εναλλακτικά, με τον ίδιο στόχο, θα μπορούσε κανείς να χρησιμοποιήσει την πολυδιάστατη γεωμετρική βαθμονόμηση ομοιοτήτων (*multidimensional scaling*) ή την πολλαπλή ανάλυση αντιστοιχιών (*multiple correspondence analysis*) για κατηγορικά στοιχεία.

Τέλος, σε συνδυασμό με την ανίχνευση πολώσεων στα ερωτήματα-δοκιμασίες, χρησιμοποιούνται άλλες μονομεταβλητές και πολυμεταβλητές μέθοδοι, όπως η ανάλυση δομικής ισοτιμίας του τεστ για διάφορες ενδοπολιτισμικές ή διαπολιτισμικές ομάδες του γενικού δείγματος, η οποία θα αναλυθεί εκτενέστερα στο κεφάλαιο της ανάλυσης παραγόντων, καθώς η δομή του τεστ εκφράζεται μέσω των παραγόντων τους οποίους –πιθανώς– εμπεριέχει.

Κριτήρια και λάθη στην τελική επιλογή ερωτημάτων

Τρία είναι τα «πρακτικά» ζητήματα που πρέπει να συζητηθούν ως προς την «τελική» επιλογή ερωτημάτων: Αρχικά, σε καμία περίπτωση δεν θα πρέπει να μας διαφύγει ότι είναι απαραίτητη η διατήρηση της εγκυρότητας αντιπροσωπευτικού περιεχομένου στην ψυχομετρική κλίμακα αξιολόγησης. Αυτό σημαίνει ότι δεν θα εξαιρέσουμε όποιο ερώτημα φαίνεται έστω και λίγο προβληματικό χωρίς πρώτα μια τέτοια πιθανότητα να περάσει από αυστηρό έλεγχο. Άλλως, θα καταλήξουμε σε πολύ μικρό τελικό αριθμό ερωτημάτων και αν παραμείνουμε εκεί τότε δεν είναι δυνατόν να αντιπροσωπεύουμε πλήρως όλες τις πιθανές πλευρές της υπό μελέτη έννοιας¹⁹.

Επίσης, κατά τη δημιουργία αντικειμενικών ψυχομετρικών εργαλείων και ειδικά εργαλείων που προσπαθούν να μετρήσουν ικανότητες, θα πρέπει να διατηρούμε ισορροπία μεταξύ «εύκολων» και «δύσκολων» ερωτημάτων. Με άλλα λόγια, είναι θεμιτό να συμπεριλάβουμε όσο το δυνατόν *ίσους* αριθμούς δύσκολων και εύκολων ερωτημάτων στην κλίμακα αυτή αξιολόγησης. Είναι προφανές ότι ο δείκτης δυσκολίας ή ευκολίας των ερωτημάτων δεν θα είναι απόλυτα ίδιος αλλά θα πρέπει αθροίζοντάς τους αλγεβρικά να προκύπτει περίπου μηδενικό αποτέλεσμα.

¹⁹ Στην κλίμακα Αξιών που κατασκεύασε ο Rokeach (1973) περιλαμβάνονται 18 ερωτήματα για τις αξίες-σκοπούς και 18 ερωτήματα για τις αξίες-μέσα. Ο ίδιος ο θεωρητικός παραδέχεται, σε σύγγραμμά του, ότι η επιλογή των ερωτημάτων αυτών είναι διαισθητική και ότι άλλοι ερευνητές μπορεί να καταλήξουν σε άλλο κατάλογο αξιών. Στην περίπτωση, ο Rokeach δεν ισχυρίζεται ότι η κλίμακα RVS διαθέτει εγκυρότητα αντιπροσωπευτικού περιεχομένου.

Αυτές οι διαπιστώσεις μας οδηγούν και πάλι στο πολύ σοβαρό ζήτημα που αφορά στην διαφοροποιητική ισχύ των ερωτημάτων. Με βάση και τόσα έχουν ήδη συζητηθεί, προς το παρόν περιοριζόμαστε να υπενθυμίσουμε ότι η διαφοροποιητική αυτή ισχύς είναι άμεσα συνδεδεμένη με τη διασπορά του κάθε ερωτήματος. Εάν ο δείκτης δυσκολίας p του ερωτήματος βρίσκεται περίπου στο 0,5 (για δίτιμες κλίμακες μέτρησης) τότε η διαφοροποιητική ισχύς είναι υψηλή ως η μέγιστη. Αυτό συμβαίνει καθώς μεγιστοποιείται η διασπορά των τιμών για το ερώτημα αυτό. Για τα ερωτήματα αυτά ισχύει ότι η διασπορά είναι ίση με $[p \otimes (1 - p)]$ και μεγιστοποιείται όταν το p είναι ίσο με 0,5. Ενδιαφέρουν λοιπόν κυρίως τα ερωτήματα εκείνα για τα οποία οι τιμές p κυμαίνονται γύρω στο 0,5. Σύμφωνα με την αρχική θεωρία της ανάλυσης ερωτημάτων εάν οι τιμές αυτές κυμαίνονται μεταξύ 0,2 και 0,8 τότε η διασπορά και η διαφοροποιητική ισχύς του ερωτήματος είναι ικανοποιητική. Πολλές φορές όμως, σύμφωνα με τις πιο πάνω διαπιστώσεις καθώς και ανάλογα με το είδος της υπό μελέτη έννοιας, τα όρια αυτά δεν είναι απόλυτα και ίσως χρειάζεται να «συρρικνωθούν» γύρω από το 0,5 (π.χ. 0,4 έως 0,6), ανάλογα και με το αντικείμενο της μέτρησης).

Ανάλυση Παραγόντων

Η «εναλλακτική» προσέγγιση στον έλεγχο ερωτημάτων είναι η ανάλυση παραγόντων ή «παραγοντική ανάλυση». Για να αποφύγουμε τη σύγχυση με τις μεθόδους ανάλυσης διακύμανσης, όπου υπάρχουν όροι όπως μονο- δι- πολυ- παραγοντική ανάλυση διακύμανσης, ο όρος ανάλυση παραγόντων είναι προτιμότερος. Η μέθοδος αυτή δημιουργήθηκε από τον *C. Spearman* στις αρχές του 20^{ου} αιώνα και εξελίχθηκε από πολλούς θεωρητικούς και στατιστικούς κατά τη διάρκεια του αιώνα, έως και σήμερα. Η βασική προσπάθεια εστιάζεται στο να ανακαλύψουμε την εσωτερική δομή, τις λανθάνουσες κυρίαρχες δομές και παραμέτρους μέσα σε ένα σύνολο μετρήσεων-τιμών που προέρχεται από ένα σύνολο δοκιμασιών-ερωτημάτων. Συνήθως το σύνολο αυτό δοκιμασιών-ερωτημάτων είναι μια ψυχομετρική κλίμακα αξιολόγησης και η διερεύνηση των θεμελιακών παραμέτρων γίνεται μέσω του πίνακα συναφειών για τις μετρήσεις-τιμές των ερωτημάτων της κλίμακας αυτής.

Η ανάλυση παραγόντων είναι μια πολύπλοκη αλγεβρική διαδικασία και επιδέχεται ιδιαίτερα πολλών εναλλακτικών προσεγγίσεων, καθώς ο υπολογιστικός αλγόριθμος επιδέχεται ιδιαίτερα εκτεταμένης παραμετροποίησης. Οι εναλλακτικές μέθοδοι που μπορεί κανείς να χρησιμοποιήσει για τη διερεύνηση του πίνακα συναφειών αντιμετωπίζουν, η καθεμία από την πλευρά της, διαφορετικά θεωρητικά ζητήματα τα οποία προκύπτουν ανάλογα με τον αριθμό των στοιχείων που διαθέτουμε, τα χαρακτηριστικά του δείγματος, το είδος των συσχετιζομένων στοιχείων, κ.τ.λ. Στα στενά πλαίσια των σημειώσεων αυτών, θα μας απασχολήσουν δύο μόνο ειδικές μέθοδοι: η ανάλυση κυρίων συνιστωσών που αποτελεί το πρώτο βήμα για την περαιτέρω ανάλυση παραγόντων και, η ορθόγωνη περιστροφή των αξόνων κατά την φάση της παραγοντικής ανάλυσης.

Προϋποθέσεις

Για να προχωρήσει κάποιος σε ανάλυση κυρίων συνιστωσών και έπειτα σε ανάλυση παραγόντων πρέπει να πληρούνται συγκεκριμένες προϋποθέσεις. Αρχικά, η κλίμακα αξιολόγησης την οποία επιχειρεί κάποιος να «παραγοντοποιήσει» πρέπει να έχει τουλάχιστον 10 δοκιμασίες-ερωτήματα. Το όριο αυτό έχει υποστηριχθεί από τους πρωτεργάτες της ανάλυσης παραγόντων, όμως αυτό δεν σημαίνει ότι με 10 ερωτήματα είμαστε σίγουροι για την ασφάλεια των αποτελεσμάτων της ανάλυσης αυτής. Ο αριθμός αυτός, απλά, αποτελεί κατώτατο όριο και ο ψυχομέτρης θα χρησιμοποιήσει στην πραγματικότητα πολύ μεγαλύτερο αριθμό ερωτημάτων, ειδικά όταν βρίσκεται στο πρώτο στάδιο δημιουργίας του ψυχομετρικού εργαλείου και χρειάζεται πολλά εναλλακτικά ερωτήματα για να επιλέξει από αυτά κάποιο τμήμα τους με αποδεκτά ψυχομετρικά χαρακτηριστικά.

Η δεύτερη προϋπόθεση αφορά στην αντιπροσωπευτικότητα του δείγματος από το οποίο θα προέλθουν τα προς παραγοντοποίηση στοιχεία. Εκτός της αντιπροσωπευτικότητας αυτής που πρέπει να επιτευχθεί μέσω των μεθόδων της τυχαίας δειγματοληψίας και καλής γνώσης των εννοιολογικών και λειτουργικών ορισμών της υπό μελέτης έννοιας, θα πρέπει να τονίσουμε ότι στην ανάλυση παραγόντων χρειάζονται τουλάχιστον 4 φορές περισσότερα άτομα από τα ερωτήματα που πρόκειται να αναλυθούν.

Επομένως, εάν προσπαθούμε μέσω ανάλυσης παραγόντων να διερευνήσουμε τις θεμελιακές παραμέτρους σε σύνολο 50 δοκιμασιών-ερωτημάτων, χρειαζόμαστε δείγμα τουλάχιστον 200 συμμετεχόντων. Πολλοί θεωρητικοί μάλιστα ανεβάζουν τους συμμετέχοντες σε 5 τουλάχιστον φορές περισσότερους των προς ανάλυση ερωτημάτων. Εννοείται πως, εάν θέλουμε να επιχειρήσουμε δύο διαφορετικές αναλύσεις παραγόντων για τμήματα δοκιμασιών-ερωτημάτων (π.χ., χωριστά για διάφορες υποκλίμακες αξιολόγησης), τότε οι ίδιοι συμμετέχοντες θα μας δώσουν στοιχεία και για τις δύο αναλύσεις παραγόντων· δεν χρειαζόμαστε, δηλαδή, μεγαλύτερο δείγμα διότι για κάθε μία ανάλυση πληρούνται οι προϋποθέσεις αναλογίας ερωτημάτων-συμμετεχόντων. Καθώς όμως, όπως συμβαίνει και με την ανάλυση ερωτημάτων, πολλές φορές υπάρχουν υπόνοιες για πιθανές διαφορές μεταξύ των δύο φύλων, θα πρέπει να επαναλάβουμε την ίδια ανάλυση παραγόντων για κάθε φύλο χωριστά και συνεπώς χρειαζόμαστε, στην περίπτωση του πιο πάνω παραδείγματος, 400 (200 άντρες και 200 γυναίκες) και όχι 200 συμμετέχοντες, ώστε για κάθε χωριστή κατά φύλο ανάλυση παραγόντων να εξακολουθούν να πληρούνται οι προϋποθέσεις αναλογίας ερωτημάτων-συμμετεχόντων. Εάν υπάρχουν επιπλέον υπόνοιες και για άλλες πιθανές διαφορές ανάλογα με τα χαρακτηριστικά του δείγματος, τότε ο αριθμός των συμμετεχόντων είναι πολλαπλάσιος του 4 αλλά και των συνδυαστικών αυτών κατηγοριών. Για παράδειγμα, εάν θέλουμε να διερευνήσουμε την παραγοντική δομή 20 ερωτημάτων που θεωρητικά μετρούν υπερκινητικότητα, τότε χρειαζόμαστε αρχικά $4 \otimes 20 = 80$ συμμετέχοντες. Γνωρίζουμε όμως ότι τα αγόρια είναι περισσότερο υπερκινητικά από τα κορίτσια, επομένως χρειαζόμαστε $2 \otimes 4 \otimes 20 = 160$ συμμετέχοντες. Εάν λάβουμε υπόψιν ότι τα παιδιά που ζουν στην Αθήνα επηρεάζονται ως προς την υπερκινητικότητα που εκδηλώνουν από διαφορετικές παραμέτρους από αυτές που επηρεάζουν αντίστοιχα τα παιδιά που ζουν σε μη-αστικές περιοχές, τότε χρειαζόμαστε τουλάχιστον $2 \otimes 2 \otimes 4 \otimes 20 = 320$ συμμετέχοντες, κ.ο.κ. Θεωρητικά, βέβαια, η τυχαία δειγματοληψία και η αντιπροσωπευτικότητα του δείγματος ισορροπεί τέτοιες πιθανές διαφορές, καθώς διασπείρει τυχαία τον πιθανό «θόρυβο» που προκαλείται από τις διαφορές αυτές. Υπάρχουν όμως πολλές περιπτώσεις στις οποίες είμαστε αναγκασμένοι να επαναλάβουμε την ανάλυση παραγόντων για διάφορες συνδυαστικές κατηγορίες και σε κάθε αναλυτικό υπολογιστικό υπόδειγμα το δείγμα μας θα πρέπει να είναι τουλάχιστον τετραπλάσιο των ερωτημάτων που χρησιμοποιούμε.

Μια τρίτη προϋπόθεση είναι ότι τα ερωτήματα που χρησιμοποιούμε θα πρέπει να αναφέρονται σε μέρη ενός γενικότερου συστήματος. Δεν μπορούμε, με άλλα λόγια, να αναλύουμε παραγοντικά ανόμοιες μεταξύ τους έννοιες. Τέτοιες έννοιες είναι αυτόνομες και η ανάλυση παραγόντων τους θα πρέπει να γίνει χωριστά, εφόσον, για κάποιο ερευνητικό λόγο, ενδιαφέρουν τον ψυχομέτρη ή τον ερευνητή. Πριν προχωρήσει κανείς σε ανάλυση παραγόντων, θα πρέπει να προβεί σε επισκόπηση του πίνακα συναφειών μεταξύ των δοκιμασιών-ερωτημάτων (ενδοσυνάφειες). Ο πίνακας αυτός θα πρέπει να περιλαμβάνει μέτριες προς υψηλές συνάφειες, οποιασδήποτε κατεύθυνσης. Στην περίπτωση αυτή ενδείκνυται η ανάλυση παραγόντων. Εάν όμως η πλειοψηφία των συναφειών είναι μηδενικές ή πολύ υψηλές (της τάξης του 0,80 και άνω) τότε ο πίνακας αυτός δεν είναι ο καταλληλότερος για ανάλυση παραγόντων καθώς οι επιμέρους μετρήσεις που προκύπτουν από τα διάφορα ερωτήματα φαίνεται να είναι ασύνδετες ή παρόμοιες έως απόλυτα όμοιες. Στην περίπτωση αυτή, για την υφιστάμενη ψυχομετρική κλίμακα αξιολόγησης δεν μπορούμε να καταλήξουμε σε θεμελιακές παραμέτρους μέσω ανάλυσης παραγόντων. Οι λόγοι για τους οποίους συμβαίνει αυτό είναι και μαθηματικοί, αλλά και ορθολογικά σκεπτομενοι καταλήγουμε στο ίδιο συμπέρασμα.

Τέλος (4^η προϋπόθεση), είναι απαραίτητο οι μετρήσεις μας να προέρχονται από κλίμακα μέτρησης τουλάχιστον ίσων διαστημάτων. Αυτό δεν σημαίνει ότι για άλλου τύπου μετρήσεις δεν μπορούμε να προβούμε σε ανάλυση παραγόντων, αλλά ότι για να χρησιμοποιήσουμε δείκτες *Pearson r* στον πίνακα συναφειών που αναλύεται είναι αναγκαίο να πληρούνται οι δύο βασικές προϋποθέσεις του στατιστικού αυτού δείκτη. Στις περιπτώσεις που η κλίμακα μέτρησης δεν είναι ίσων διαστημάτων ή ίσων λόγων, υπολογίζονται άλλοι δείκτες συνάφειας (οι κατάλληλοι ανάλογα με την περίπτωση) και η ανάλυση παραγόντων εφαρμόζεται στους δείκτες αυτούς. Στην περίπτωση που η κλίμακα μέτρησης είναι δίτιμη (0 & 1) στην πραγματικότητα ο πίνακας συναφειών που αναλύεται περιέχει δείκτες συνάφειας Φ και όχι *Pearson r*. Αριθμητικά, οι δείκτες αυτοί ταυτίζονται, αλλά χρειάζεται ιδιαίτερη προσοχή, ώστε τα πρόσημα των δεικτών να μην είναι εσφαλμένα λόγω πιθανής αντίστροφης κωδικοποίησης. Σε άλλες περιπτώσεις μπορούν να χρησιμοποιηθούν ο δείκτης τετραχορικής συνάφειας, ο δείκτης *Spearman Rho* αυτούσιος ή με παραλλαγή της υπολογιστικής διαδικασίας, ο δείκτης *Kendall Tau*, ο δείκτης η και άλλοι δείκτες συνάφειας.

Μαθηματική βάση της ανάλυσης παραγόντων

Από θεωρητική άποψη η ανάλυση παραγόντων δεν είναι τίποτε περισσότερο παρά ένας γραμμικός συνδυασμός μετρήσεων για τις οποίες, με βάση τα υπάρχοντα στοιχεία, υπολογίζονται συντελεστές, οι οποίοι και, αρχικά, καθορίζουν την κάθε κύρια θεμελιακή παράμετρο (συνιστώσες, άξονες) και μετά από περιστροφή, τον κάθε παράγοντα (μετατροπή των συνιστωσών-αξόνων). Ο γραμμικός αυτός συνδυασμός έχει ως εξής:

$$F_i = \sum_{j=1}^v W_{ij}X_j = W_{i1}X_1 + W_{i2}X_2 + \dots + W_{iv}X_v \quad \text{όπου, } j \text{ είναι ο}$$

αριθμός των μετρήσεων που συνδυάζονται γραμμικά στον παράγοντα αυτόν, X οι αρχικές μετρήσεις και όπου $i = 1, 2, 3, \dots, k$ παράγοντες. Οι συντελεστές για κάθε μέτρηση είναι οι δείκτες W .

Ο σκοπός της παρουσίασης δεν είναι βέβαια η πλήρης κάλυψη της θεωρίας της ανάλυσης παραγόντων. Όμως, η ανάλυση βασίζεται σε μαθηματικές μετατροπές και πράξεις πινάκων γραμμικής άλγεβρας, διαδικασίες που καλύπτονται από την επιστήμη των Μαθηματικών και είναι αναγκαίο να γνωρίζουμε μερικά από τα στοιχεία τους, ώστε να έχουμε γνώση των εσωτερικών διαδικασιών. Έτσι, θα πρέπει πρώτα να παρουσιασθούν μερικές από τις βασικές έννοιες της ανάλυσης παραγόντων καθώς και η βασική λογική της, ώστε ο αναγνώστης να είναι σε θέση να εκτιμήσει και να συνειδητοποιήσει μερικά από τα πλεονεκτήματα αλλά και μειονεκτήματα της μεθόδου.

Δείκτες συνάφειας

Όταν έχουμε συλλέξει στοιχεία για, έστω, 10 ερωτήματα προσωπικότητας, οι δείκτες συνάφειας που προκύπτουν μεταξύ των ερωτημάτων αυτών (στην περίπτωση, $9+8+7+6+5+4+3+2+1 = 45$ ή $9 \otimes 10 \div 2 = 45$ ενδοσυνάφειες) είναι δύσκολο να ερμηνευθούν στο σύνολό τους μόνο και μόνο με επισκόπηση του πίνακα συναφειών. Ο Πίνακας της σελ. 78 αποτελεί παράδειγμα, αν και όχι χαρακτηριστικό, καθώς περιέχει «ιδανικούς» δείκτες. Στην ιδανική αυτή περίπτωση, δύο «ομάδες» ενδοσυναφειών είναι προφανείς κατά την οπτική επισκόπηση του πίνακα (στον πίνακα αυτόν άλλοι δείκτες είναι υπογραμμισμένοι και άλλοι τονίζονται με έντονους χαρακτήρες). Σπανιότατα, όμως, προκύπτουν τέτοιου είδους

ενδοσυνάφειες. Ο Πίνακας συναφειών είναι συμμετρικός, όπως γνωρίζουμε, και για τον λόγο αυτόν παρουσιάζεται ως τριγωνικός πίνακας. Η διαγώνιος περιέχει τις ενδοσυνάφειες των μετρήσεων με τον εαυτό τους και για τον λόγο αυτόν οι δείκτες αυτοί είναι παντού μοναδιαίοι. Για τις μετρήσεις, επομένως, M1 έως M10, έστω ότι έχουν προκύψει οι ακόλουθοι δείκτες:

Υποθετικοί δείκτες συνάφειας μεταξύ 10 μετρήσεων

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
M1	1,00									
M2	0,45	1,00								
M3	0,35	0,34	1,00							
M4	0,02	0,07	0,16	1,00						
M5	0,64	0,52	0,47	0,12	1,00					
M6	0,48	0,39	0,46	0,14	0,39	1,00				
M7	0,49	0,37	0,69	0,05	0,42	0,45	1,00			
M8	0,10	-0,01	-0,06	0,68	-0,10	-0,14	0,61	1,00		
M9	0,08	0,05	0,09	0,46	0,03	-0,01	0,59	0,40	1,00	
M10	0,13	0,04	0,15	0,55	0,01	0,07	0,48	0,38	0,51	1,00

Είναι σημαντικό να τονίσουμε ότι για τους δείκτες αυτούς δεν ενδιαφέρει η στατιστική σημαντικότητα. Αντίθετα, μας ενδιαφέρει ιδιαίτερα ο βαθμός συνάφειας καθώς και η κατεύθυνση. Πρέπει επίσης να τονίσουμε, ότι και στην περίπτωση αυτή (καθώς οι συνήθως χρησιμοποιούμενοι δείκτες είναι δείκτες *Pearson r*) δεχόμαστε τις δύο στατιστικές προϋποθέσεις που είναι αναγκαίες για την περαιτέρω χρήση των δεικτών αυτών. Δεχόμαστε, δηλαδή, εκ των προτέρων ότι οι μετρήσεις (M1 έως M10) είναι αριθμητικές και ότι η σχέση μεταξύ των μετρήσεων είναι ευθύγραμμη και όχι καμπυλόγραμμη. Παρατηρούμε ότι οι μετρήσεις M1, M2, M3, M5, M6 και M7 φαίνεται να παρουσιάζουν αρκετά υψηλές ενδοσυνάφειες και ότι, αντίστοιχα, οι μετρήσεις M4, M8, M9 και M10 παρουσιάζουν επίσης υψηλές ενδοσυνάφειες μεταξύ τους, ενώ μεταξύ τους οι δύο αυτές «ομάδες» δεν παρουσιάζουν ενδοσυνάφειες. Σύμφωνα με τις παρατηρήσεις αυτές, είναι πιθανόν τα ερωτήματα του ψυχομετρικού μας εργαλείου να εμπεριέχουν δύο θεμελιακές παραμέτρους και αυτό θα προσπαθήσουμε να διαλευκάνουμε χρησιμοποιώντας την ανάλυση παραγόντων.

Βασική Ορολογία ανάλυσης παραγόντων

Η στατιστική επεξεργασία των πληροφοριών αυτών οδηγεί σε διάφορα μεγέθη-δείκτες οι οποίοι και είναι απαραίτητοι για την περαιτέρω ερμηνεία των ευρημάτων. Μερικοί από τους δείκτες αυτούς είναι οι ακόλουθοι:

α) *Kaiser-Meyer-Olkin coefficient of sampling adequacy*. Ο δείκτης αυτός (δείκτης καταλληλότητας του δείγματος για ανάλυση παραγόντων) κυμαίνεται μεταξύ του 0 και του 1 και όσο πλησιάζει τη μονάδα, τόσο κατάλληλο είναι το δείγμα μας για τη μέθοδο που επιχειρούμε. Αν ο δείκτης αυτό είναι σχετικά μικρός ($<0,7$), τότε η ανάλυση παραγόντων μπορεί να μην είναι και τόσο καλή ιδέα για την περίπτωση. Το κριτήριο αυτό συνδυάζεται με πληροφορίες που προέχρονται από την ορίζουσα (*determinant*) του πίνακα συναφειών.

β) *Bartlett's test of sphericity*. Ο δείκτης αυτός αναφέρεται στην ομοιογένεια του δείγματος και ελέγχει την πιθανότητα μοναδιαίας συμμετρίας (*compound symmetry*) στον πίνακα διασποράς-συνδιασποράς των μετρήσεων (*variance-covariance matrix*), δηλαδή ελέγχει το αν στη διαγώνιο του πίνακα υπάρχουν μονάδες και παντού εκτός διαγωνίου μηδενικά. Ο δείκτης κρίνεται για τη στατιστική του σημαντικότητα μέσω του στατιστικού δείκτη F. Αν ο δείκτης αυτός αποδειχθεί στατιστικά σημαντικός, τότε οι τιμές εκτός διαγωνίου διαφέρουν από το μηδέν και υφίσταται σφαιρικότητα στα στοιχεία που διαθέτουμε, δεν τίθεται, δηλαδή, ζήτημα ανυπαρξίας συναφειών στα δεδομένα μας. Το κριτήριο αυτό είναι διαβόητο για τη στατιστική σημαντικότητά του, καθώς και πολύ μικρές συνδιασπορές (που έπειτα καταλήγουν σε συνάφειες) είναι δυνατό να είναι στατιστικά σημαντικές για ένα σχετικά μεγάλο πλήθος ατόμων. Έτσι το κριτήριο αυτό συνήθως είναι στατιστικά σημαντικό, δείχνει δηλαδή την ύπαρξη συναφειών που λογικά μπορούν να αναλυθούν, όμως δεν εξασφαλίζει ότι το μέγεθος των αναλυόμενων συναφειών είναι αρκετά ικανοποιητικό.

γ) *Eigenvalues*. Στο σημείο αυτό έχει αρχίσει η διαδικασία μετατροπής των πινάκων καθώς και οι πράξεις πινάκων. Η διαγωνοποίηση του πίνακα συναφειών απολήγει μέσω τελεστών, που επίσης υπολογίζονται με βάση τον ίδιο πίνακα, σε *Ιδιοτιμές* οι οποίες αντιστοιχούν στη διασπορά τιμών που συνδέεται με κάθε θεμελιακή παράμετρο. Για κάθε μέτρηση που παίρνει μέρος στην ανάλυση παραγόντων, υπολογίζεται και η αντίστοιχη ιδιοτιμή. Πρέπει να διευκρινίσουμε από το σημείο αυτό και έπειτα αρχίζει η διαδικασία ανεύρεσης των θεμελιακών παραμέτρων (κύριες συνιστώσες, κύριοι άξονες). Θα μιλήσουμε μόνο για την ανάλυση κυρίων συνιστωσών (*Principal Components Analysis*) η οποία επιχειρεί και να ομαδοποιήσει τα στοιχεία που διαθέτουμε με βάση τις θεμελιακές παραμέτρους. Για κάθε ιδιοτιμή υπολογίζεται και το ακριβές ποσοστό της διασποράς των

τιμών που συνδέεται με την κάθε θεμελιακή παράμετρο. Στο ολοκληρωμένο παράδειγμα ανάλυσης παραγόντων που θα ακολουθήσει θα επανέλθουμε στην έννοια της ιδιοτιμής.

δ) *Factor loading*. Η φόρτιση²⁰ (ή φορτίο) της κάθε αρχικής μέτρησης (*item*) στην κάθε θεμελιακή παράμετρο που ανιχνεύθηκε μέσω της ανάλυσης κυρίων συνιστωσών υπολογίζεται ως δείκτης συνάφειας της μέτρησης αυτής με την κάθε κύρια συνιστώσα. Η ίδια διαδικασία επαναλαμβάνεται και στο δεύτερο στάδιο της ανάλυσης, μετά την περιστροφή των θεμελιακών παραμέτρων, όπου πλέον η συνάφεια εκφράζει τη σχέση της κάθε μέτρησης με τον κάθε παράγοντα. Θυμίζουμε ότι οι κύριες συνιστώσες καθώς και οι παράγοντες ορίζονται μαθηματικά ως οι γραμμικοί συνδυασμοί (μέσω συντελεστών) των μετρήσεων που μετέχουν στην κύρια αυτή συνιστώσα ή στον παράγοντα αυτόν και οι συνάφειες της κάθε μέτρησης εκφράζουν τη συσχέτιση με τους γραμμικούς αυτούς συνδυασμούς.

ε) *Factor rotation*. Η περιστροφή των αξόνων κατά την ανάλυση παραγόντων είναι και το σπουδαιότερο διερευνητικό στάδιο, αν και οι θεμελιακές παράμετροι έχουν ήδη προκύψει, καθώς με αλγοριθμικούς επανυπολογισμούς (*iterations*) προσπαθούμε να μεγιστοποιήσουμε τη συνάφεια (φόρτιση) της κάθε μέτρησης με έναν μόνο από τους υπάρχοντες γραμμικούς συνδυασμούς. Ταυτόχρονα ελαχιστοποιούμε τα αντίστοιχα φορτία με τους υπόλοιπους γραμμικούς συνδυασμούς-παράγοντες. Η προσπάθεια αυτή μπορεί να γίνει με πολλούς τρόπους, αλλά εάν θέλουμε να διατηρήσουμε την ανεξαρτησία των παραγόντων (δηλαδή, αν θέλουμε να θεωρήσουμε ότι η διαμόρφωση του ενός δεν επηρεάζει τους άλλους παράγοντες) τότε χρησιμοποιούμε την ορθόγωνη περιστροφή των αξόνων (*Orthogonal rotation*). Η απόφαση αυτή εξαρτάται και από τα ίδια τα στοιχεία που έχουμε συλλέξει, από τη θεωρητική βάση που διέπει την όλη κατασκευή του ψυχομετρικού εργαλείου, καθώς και από άλλες παραμέτρους.

στ) *Factor scores*. Μετά τον υπολογισμό των ιδιοτιμών και των φορτίων κάθε μέτρησης στις κύριες συνιστώσες και στους παράγοντες, υπολογίζονται μέσω μετατροπής των αρχικών τιμών και εφαρμογής της συνάρτησης που εκφράζει μαθηματικά τον κάθε παράγοντα οι χαρακτηριστικές «παραγοντικές» τιμές του κάθε συμμετέχοντα για κάθε παράγοντα. Επομένως, εάν μια ανάλυση παραγόντων καταλήξει σε 4 παράγοντες, για κάθε

²⁰ Βλ. Ευκλείδη, Α & Κάντας, Α. (1998) Γλωσσάρι, *Ψυχολογία*, 5(1), 84-88.

συμμετέχοντα θα υπολογισθούν 4 παραγοντικές τιμές, οι οποίες και μπορούν να χρησιμοποιηθούν σε περαιτέρω διαδικασίες ανάλυσης.

Ως προς το θέμα της οικονομίας (*Parsimony*), μέσω όλων των πιο πάνω διαδικασιών επιτυγχάνεται συνοπτική περιγραφή των αρχικών δεικτών συνάφειας με λίγες θεμελιακές παραμέτρους που προκύπτουν από αυτούς. Εάν η ανάλυση οδηγήσει σε επιτυχή περιστροφή των θεμελιακών παραμέτρων, τότε πιθανόν να επιτευχθεί και εννοιολογική σαφήνεια (*conceptual meaningfulness*).

Αποτελέσματα ανάλυσης κυρίων συνιστωσών και αποτελέσματα ανάλυσης παραγόντων

Στο υποθετικό μας παράδειγμα, μετά τον υπολογισμό των διαφόρων δεικτών μπορούμε να κατασκευάσουμε τον ακόλουθο Πίνακα, ο οποίος και αναφέρεται στα αποτελέσματα της ανάλυσης κυρίων συνιστωσών.

Αποτελέσματα ανάλυσης κυρίων συνιστωσών (αρχικοί δείκτες)

Ορίζουσα του πίνακα συναφειών = 0,0097066

Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0,68225

Bartlett Test of Sphericity = 110,46636, Significance = 0,00000

Extraction 1 for analysis 1, Principal Components Analysis (PC)

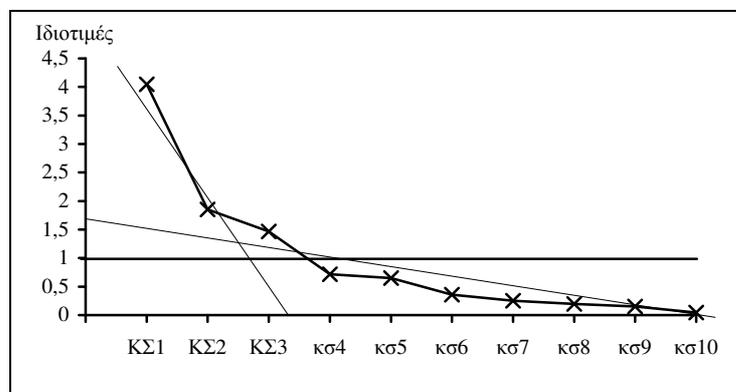
Initial Statistics:

Μέτρηση	Communality	Παράγων	Ιδιοτιμή	% αποδιδόμενης διασποράς	Αθροιστικό % αποδιδόμενης διασποράς
ME1	1,0000	1	4,04757	40,5	40,5
ME2	1,0000	2	1,85512	18,16	59,0
ME3	1,0000	3	1,46433	11,4	70,4
ME4	1,0000	4	0,71456	7,2	77,6
ME5	1,0000	5	0,64672	6,2	83,8
ME6	1,0000	6	0,35987	4,2	88,0
ME7	1,0000	7	0,24987	4,0	92,0
ME8	1,0000	8	0,19645	3,8	95,7
ME9	1,0000	9	0,15328	2,8	98,5
M10	1,0000	10	0,04322	1,5	100,0

PC extracted 3 factors.

Από τον Πίνακα αυτόν συνάγονται διάφορα συμπεράσματα: Πρώτον, οι ιδιοτιμές μας πληροφορούν για την ύπαρξη τριών κυρίων συνιστωσών. Έχουν όμως και οι τρεις κύριες συνιστώσες κάποιο νόημα, είναι δηλαδή και οι τρεις θεμελιακές παράμετροι χρήσιμες και ισχυρές; Παρατηρούμε ότι για την πρώτη κύρια συνιστώσα

το ποσοστό της αποδιδόμενης στη συσχέτιση διασποράς των τιμών, δηλαδή της διασποράς των τιμών που οφείλεται στον γραμμικό συνδυασμό μετρήσεων που αποτελούν τη συνιστώσα αυτή είναι περίπου 40%. Για τη δεύτερη κύρια συνιστώσα το ποσοστό αυτό είναι περίπου 18% και για την τρίτη 11,4%. Όταν στην ανάλυση κυρίων συνιστωσών, η πρώτη από αυτές συνοδεύεται από υψηλή ιδιοτιμή τότε μπορεί να συμβαίνουν τρία πράγματα: ότι η συνιστώσα αυτή εκφράζει τη συνολικότητα της ψυχομετρικής μας κλίμακας, το γεγονός δηλαδή ότι είναι μια κλίμακα αξιολόγησης, ότι μπορεί να υπάρχουν επιδράσεις της τάσης που έχουν οι συμμετέχοντες γενικά να συμφωνούν με όλα ή να διαφωνούν με όλα (*acquiescence effects*), μια τάση που εντάσσεται στα πλαίσια πόλωσης των απαντήσεων λόγω γενικότερων επιδράσεων συμμετεχόντων (*participant effects*) ή ότι η ψυχομετρική κλίμακα αξιολόγησης είναι μονοδιάστατη, δηλαδή περιλαμβάνει μόνο μία θεμελιακή παράμετρο. Γνωρίζοντας τα παραπάνω και προσπαθώντας να εκτιμήσουμε πόσες κύριες συνιστώσες έχουν κάποια σημασία για τα ερωτήματά μας, προσωρούμε στο *Cattell Scree test*, το οποίο δεν είναι παρά μια απεικόνιση των ιδιοτιμών (ή των ποσοστών της εξηγούμενης διασποράς τιμών). Το Σχήμα που ακολουθεί παρουσιάζει το τεστ αυτό.



Οι «κύριες συνιστώσες» 4 έως 10 έχουν ιδιοτιμές μικρότερες της μονάδας και στην ουσία δεν αποτελούν κύριες συνιστώσες (*κριτήριο Kaiser-Guttman*). Το «πρόβλημα» αφορά στην κύρια συνιστώσα 3. Εάν παραμείνει, τότε, με 10 αρχικά ερωτήματα, καταλήγουμε σε 3 θεμελιακές παραμέτρους που μάλλον είναι πολλές, καθώς δεν θα πρέπει να λησμονούμε και το θέμα της οικονομίας. Από την άλλη, διαπιστώνουμε στο *scree test* ότι από την πρώτη και δεύτερη κύρια συνιστώσα μπορούμε να σύρουμε μια ευθεία η οποία είναι τελείως διαφορετική από μια άλλη ευθεία που μπορούμε να σύρουμε για τις

υπόλοιπες κύριες συνιστώσες. Αυτή η διαφοροποίηση ορίζει και το σημείο «τομής» (όπως λέγεται σε άλλο αναλυτικό υπόδειγμα, έναν «αγκώνα»). Στο σημείο αυτό, δηλαδή, φαίνεται να αποκόπτονται οι όχι και τόσο σημαντικές κύριες συνιστώσες και στην περίπτωση, αποφασίζουμε ότι οι κύριες συνιστώσες που ενδιαφέρουν είναι δύο²¹. Θα πρέπει να τονίσουμε ότι οι δύο αυτές κύριες συνιστώσες δεν ταυτίζονται με τις μετρήσεις M1 και M2 (βλ. και Πίνακα 5.) *Καμία* κύρια συνιστώσα δεν ταυτίζεται με τις αντίστοιχες στον Πίνακα μετρήσεις. Χρησιμοποιούμε τις μετρήσεις αυτές μόνο και μόνο για να περιγράψουμε το πλήθος των κυρίων συνιστωσών, έχοντας *μόνο* ως σημείο αναφοράς τις συγκεκριμένες μετρήσεις.

Η επόμενη πληροφορία που μας δίνει η ανάλυση κυρίων συνιστωσών είναι το συνολικό ποσοστό της διασποράς που οφείλεται στις κύριες συνιστώσες που έχουμε αποφασίσει ότι υπάρχουν στα ερωτήματά μας. Στην περίπτωση, το ποσοστό αυτό είναι 59%. Το ποσοστό αυτό σημαίνει από το σύνολο των υπαρχόντων στις μετρήσεις μας πληροφοριών, το 59% αποδίδεται στην ύπαρξη των συγκεκριμένων κυρίων συνιστωσών και όχι σε άλλες τυχαίες ή συστηματικές πηγές. Επομένως, η τυχαιότητα που συμπεριλαμβάνει και τους συστηματικούς παράγοντες των ατομικών διαφορών που οφείλονται σε άλλες παραμέτρους, διαφορετικές από εκείνη που το ψυχομετρικό μας εργαλείο επιχειρεί να μετρήσει, αλλά και τη διασπορά λάθους, το σφάλμα δηλαδή της τυχαίας δειγματοληψίας, είναι 41%. Εάν, βέβαια, ο ψυχομέτρης καταφέρει να αποδώσει στο ψυχομετρικό του εργαλείο το 59% της διασποράς των τιμών ευχαριστεί τις καλές του μοίρες, καθώς κάτι τέτοιο δεν είναι καθόλου εύκολο.

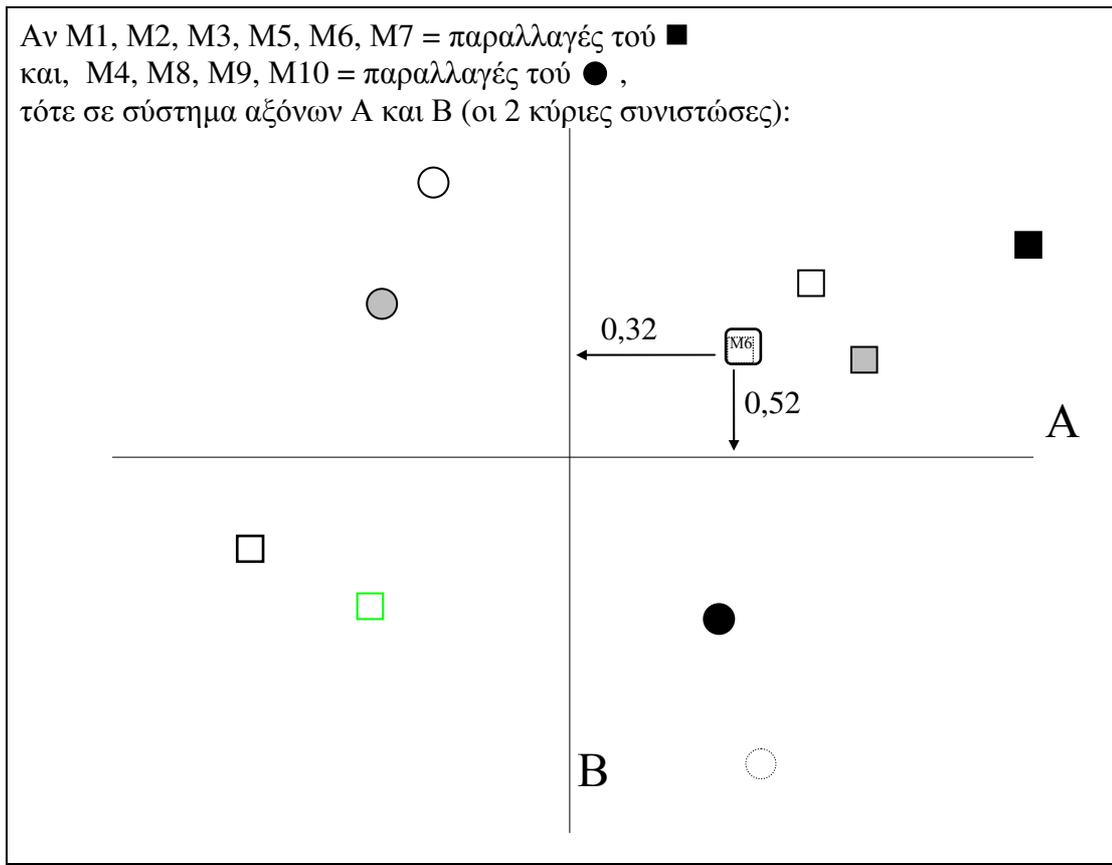
Το επόμενο στάδιο της ανάλυσης κυρίων συνιστωσών αναφέρεται στον υπολογισμό των φορτίσεων της κάθε μέτρησης σε κάθε μία κύρια συνιστώσα. Οι φορτίσεις αυτά κυμαίνονται μεταξύ του -1,00 και του +1,00, καθώς είναι, επίσης, δείκτες συνάφειας. Ο Πίνακας που ακολουθεί παρουσιάζει τις φορτίσεις των μετρήσεων στις δύο κύριες συνιστώσες.

²¹ Βλ. **Tabachnick & Fidell** (1989) *Using Multivariate Statistics*, σελ. 635.

Αποτελέσματα ανάλυσης κυρίων συνιστωσών

Μέτρηση	Ιδιοτιμή	Κ.Συνιστώσα 1	Κ. Συνιστώσα 2
M1	4,04757	0,58	0,10
M2	1,85512	0,61	0,37
M3	1,46433	0,44	-0,41
M4	0,71456	-0,35	0,48
M5	0,64672	0,36	0,36
M6	0,35987	0,52	0,32
M7	0,24987	0,57	0,27
M8	0,19645	-0,23	0,47
M9	0,15328	-0,31	0,53
M10	0,04322	0,30	0,60

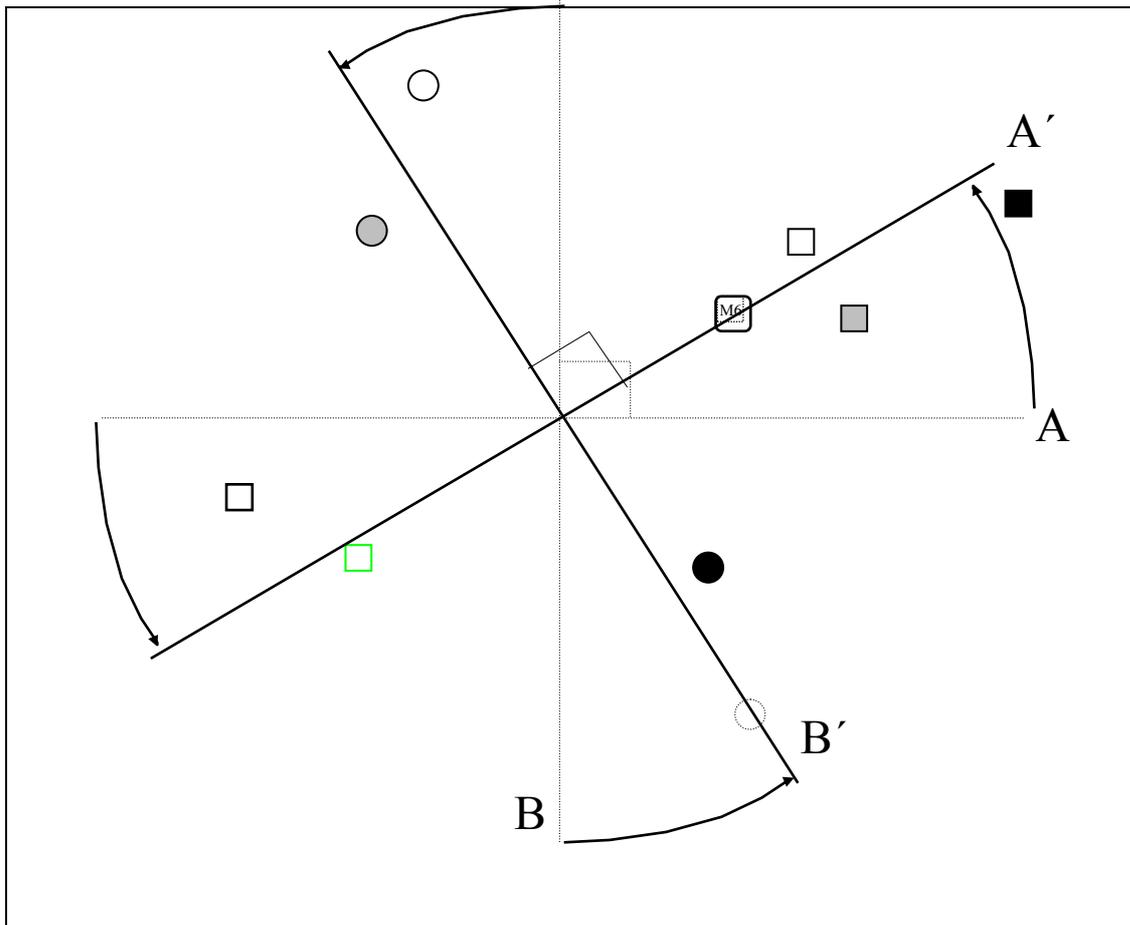
Παρατηρούμε ότι για την πρώτη κύρια συνιστώσα όλες σχεδόν οι μετρήσεις παρουσιάζουν φορτίσεις μεγαλύτερες του 0,30. Συνήθως, στην ανάλυση παραγόντων γενικά δεχόμαστε ότι η φόρτιση είναι ενδιαφέρουσα εάν είναι μεγαλύτερη από την τιμή αυτή, ή αν είμαστε πιο αυστηροί από την τιμή 0,40. Υπάρχουν συγκεκριμένες διαδικασίες με βάση τις οποίες είναι δυνατό να καθορίσουμε επακριβώς το σημείο αυτό «τομής», αλλά η πείρα έχει δείξει ότι σε μια αποδεκτή ανάλυση παραγόντων, που πληροί τα κριτήρια και τις προϋποθέσεις, ένα αποδοτικό σημείο τομής είναι το 0,40 (απόλυτη τιμή). Στην περίπτωση, όλες οι μετρήσεις φαίνεται να είναι άμεσα συνδεδεμένες με την πρώτη κύρια συνιστώσα. Αυτή η διαπίστωση δείχνει και ότι αυτή η κύρια συνιστώσα είναι η ίδια η ύπαρξη της ψυχομετρικής κλίμακας των 10 ερωτημάτων, αλλά ταυτόχρονα δείχνει και την πιθανή παρείσφρυση του εξωγενούς παράγοντα του *acquiescence*. Για τη δεύτερη κύρια συνιστώσα 8 από τα 10 ερωτήματα φαίνεται να είναι συναφή. Η σπουδαιότερη όμως παρατήρηση είναι ότι 7 ερωτήματα (τα 2, 3, 4, 5, 6, 9 και 10) έχουν ταυτόχρονα μέτρια προς υψηλή φόρτιση και στις δύο κύριες συνιστώσες. Αυτό συμβαίνει, καθώς στο πρώτο αυτό στάδιο οι συνάψεις κάθε μέτρησης με τους γραμμικούς συνδυασμούς των θεμελιακών παραμέτρων δεν έχουν ακόμη μεγιστοποιηθεί. Παρουσιάζοντας σχηματικά τις φορτίσεις αυτές για το στάδιο της ανάλυσης κυρίων συνιστωσών θα καταλήγαμε περίπου στα ακόλουθα **:



** Το Σχήμα δεν αντιπροσωπεύει τον Πίνακα συναφειών. Παρουσιάζεται μόνον χάριν παραδείγματος.

Στο Σχήμα παρατηρούμε ότι η Μέτρηση 6 έχει σχετικά υψηλή φόρτιση στην Κύρια συνιστώσα A και μέτρια φόρτιση στην κύρια συνιστώσα B. Αυτή η εικόνα δεν μας δια φωτίζει ως προς την εννοιολογική οντότητα της μέτρησης M6. Το ίδιο συμβαίνει και για άλλες μετρήσεις και έτσι, ενώ γνωρίζουμε ότι υπάρχουν δύο θεμελιακές παράμετροι στα 10 ερωτήματα του ψυχομετρικού εργαλείου, δεν γνωρίζουμε τι ακριβώς αντιπροσωπεύουν οι παράμετροι αυτές.

Για να διευκρινισθεί η πιο πάνω εικόνα προχωρούμε σε περιστροφή των αξόνων (στην περίπτωση ορθόγων) επιχειρώντας να «ταιριάξουμε» τις συνιστώσες μας, όπως παρουσιάζονται στο σύστημα των δύο αξόνων, στη διάταξη των μετρήσεων, δηλαδή να ανακαλύψουμε την καλύτερη δυνατή λύση ως προς τα στοιχεία μας που θα μεγιστοποιήσει τις φορτίσεις και θα δια φωτίσει εννοιολογικά τις θεμελιακές παραμέτρους. Μετά από την ορθόγωνη περιστροφή των αξόνων, η κατάσταση διαγράφεται όπως στο Σχήμα** που ακολουθεί.



** Το Σχήμα δεν αντιπροσωπεύει τον Πίνακα συναφειών. Παρουσιάζεται μόνον χάριν παραδείγματος.

Η περιστροφή των αξόνων δεν αποδίδει με την πρώτη προσπάθεια το προσδοκώμενο αποτέλεσμα. Χρειάζονται διαδοχικές περιστροφές, οι οποίες γίνονται με βάση τριγωνομετρικούς δείκτες (*factor cosines*) για την ανεύρεση της τελικής, καλύτερης δυνατής λύσης, η οποία μπορεί και να μην προκύψει. Στον Πίνακα που ακολουθεί παρουσιάζονται οι φορτίσεις κάθε μέτρησης με τους παράγοντες που έχουν προκύψει από την ανάλυση κυρίων συνιστωσών και έχουν διευκρινισθεί μετά την περιστροφή των θεμελιακών παραμέτρων (αξόνων). Παρατηρούμε ότι οι φορτίσεις αυτές έχουν μεγιστοποιηθεί για τη σχέση με τον ένα παράγοντα και ταυτόχρονα έχουν ελαχιστοποιηθεί για τη σχέση των ίδιων μετρήσεων με τον άλλο παράγοντα.

Αποτελέσματα Ανάλυσης Παραγόντων με ορθόγωνη περιστροφή αξόνων

Μέτρηση	Ιδιοτιμή	Παράγοντας 1	Παράγοντας 2
M1	14,85	0,70	0,05
M2	7,6	0,65	0,13
M3	-	0,59	-0,25
M4	-	-0,12	0,62
M5	-	0,48	0,24
M6	-	0,85	0,11
M7	-	0,69	0,11
M8	-	-0,13	0,57
M9	-	-0,17	0,59
M10	-	0,17	0,68

Έτσι, μπορούμε να καταλήξουμε στο συμπέρασμα ότι οι μετρήσεις M1, M2, M3, M5, M6 και M7 αποτελούν τον ένα παράγοντα ενώ οι μετρήσεις M4, M8, M9 και M10 τον άλλο παράγοντα. Απομένει η επισκόπηση των εννοιών που τα αντίστοιχα ερωτήματα καλύπτουν για να καταλήξουμε και στην εννοιολογική ταυτότητα του κάθε παράγοντα. Αν για παράδειγμα τα ερωτήματα M1, M2, M3, M5, M6 και M7 αναφέρονται στην επιθετική συμπεριφορά ενώ τα M4, M8, M9 και M10 στην εξωστρέφεια, τότε ο πρώτος παράγοντας είναι η επιθετικότητα και ο δεύτερος η εξωστρέφεια. Κατά συνέπεια, το ψυχομετρικό εργαλείο που έχουμε δημιουργήσει έχει δύο διαστάσεις και μετρά την επιθετικότητα και εξωστρέφεια του εξεταζομένου.

Προσπαθήσαμε, επί τη βάση ενός «πλαστού» παραδείγματος να προσεγγίσουμε μερικές από τις έννοιες καθώς και τη βασική λογική της ανάλυσης παραγόντων. Ο προσεκτικός αναγνώστης θα διαπίστωσε ότι δεν αναφερθήκαμε καθόλου στην έννοια του *communality* η οποία αποτελεί και έχει θεμελιώδη-καθοριστική ισχύ για την ανάλυση παραγόντων. Η προσέγγιση όμως της έννοιας αυτής δεν είναι απαραίτητη επί του παρόντος, ούτε και μπορεί να γίνει αντικείμενο εκτενούς συζήτησης στα στενά πλαίσια των ανά χείρας σημειώσεων.

Κλείνοντας το κεφάλαιο της ανάλυσης παραγόντων και το γενικότερο πλαίσιο του ελέγχου των ερωτημάτων μιας ψυχομετρικής διαδικασίας, θα πρέπει να επισημάνουμε ότι οι μέθοδοι στις οποίες αναφερθήκαμε δεν είναι οι μοναδικές. Επίσης, θα πρέπει να γνωρίζουμε ότι οι ίδιες μέθοδοι χρησιμοποιούνται και για καθαρά ερευνητικούς σκοπούς, ασχέτως με την δημιουργία, προσαρμογή ή στάθμιση

κάποιου ψυχομετρικού εργαλείου. Τέλος, πρέπει επίσης να τονίσουμε ότι δεν αποτελούν πανάκεια, ούτε και μας εξασφαλίζουν ως προς την καταλληλότητα ενός ψυχομετρικού μέσου. Η βάση της δημιουργίας ενός ψυχομετρικού εργαλείου αποτελείται από πολλούς, προς επίτευξη, στόχους: την αξιοπιστία, την εγκυρότητα, τη στερεή θεωρητική βάση, τη χρησιμότητα και την πρωτοτυπία, την πρακτική αναγκαιότητα, την αποφυγή πολώσεων, τον έλεγχο ερωτημάτων και πολλούς άλλους στόχους οι οποίοι πρέπει να αντιμετωπισθούν στο κατάλληλο στάδιο δημιουργίας του μέσου και να στεφθούν από επιτυχία για να μπορέσουμε να προχωρήσουμε σε επόμενα στάδια.

Μέρος 7. Λανθάνοντα χαρακτηριστικά (*latent traits* - 'θ')

Θεωρία των Λανθανόντων Χαρακτηριστικών (Item Response Theory)

Αν έχουμε στη διάθεσή μας τις μεθόδους ελέγχου ερωτημάτων, σε πολύ μεγάλο ποσοστό το οφείλουμε στην ανάπτυξη των αντίστοιχων θεωριών των λανθανόντων χαρακτηριστικών (ή θεωριών υπολογισμού των χαρακτηριστικών). Τα θεωρητικά αυτά πλαίσια, με τις παραλλαγές και τις διαφορές χρησιμότητές τους, μας έχουν επιτρέψει να συστηματοποιήσουμε, τουλάχιστον στα πλαίσια της ψυχομετρικής προσέγγισης των ανθρώπινων ικανοτήτων, τον τρόπο δημιουργίας των αντίστοιχων ψυχομετρικών εργαλείων. Αυτό, βέβαια, δεν σημαίνει ότι για όλους τους άλλους τομείς που ενδιαφέρουν την ψυχομετρία τα θεωρητικά αυτά πλαίσια δεν ισχύουν. Το αντίθετο, χρησιμοποιούνται ευρύτατα με κάποιες διαφοροποιήσεις ανάλογα με την περίπτωση. Η θεωρητική αυτή βάση αναπτύχθηκε στα πλαίσια της μέτρησης της νοημοσύνης και των ικανοτήτων και στα πλαίσια αυτά θα επικεντρωθεί και η παρουσίαση που ακολουθεί.

Το πρώτο αξίωμα που διέπει τη θεωρία των λανθανόντων χαρακτηριστικών είναι ότι το λανθάνον χαρακτηριστικό είναι στατιστική οντότητα και όχι απλά ένα ψυχολογικό χαρακτηριστικό. Η πολυπλοκότητα που εισάγει το αξίωμα αυτό δεν θα συζητηθεί επί του παρόντος, όμως θα είναι καλό να αναφέρουμε ότι στην πολυπλοκότητα αυτήν στηρίζονται όλες οι προσπάθειες που γίνονται τις τελευταίες δεκαετίες για την δημιουργία πολυδιάστατων ψυχομετρικών εργαλείων. Αρκεί,

πάντως, να αναφέρουμε ότι η στατιστική αυτή οντότητα εκφράζει μαθηματικά ένα σύνολο χαρακτηριστικών, γνωρισμάτων, ποιοτήτων και εννοιών που εμπεριέχονται σε ένα γενικότερο πλαίσιο, όπως αυτό των ικανοτήτων. Όταν θα χρησιμοποιήσουμε λοιπόν την έννοια της ικανότητας ως λανθάνοντος χαρακτηριστικού, δεν θα εννοούμε μία συγκεκριμένη ικανότητα, αλλά ούτε και το σύνολο των ικανοτήτων που μπορεί να παρουσιάζει ένας άνθρωπος, αλλά θα εννοούμε τη συνάρτηση εκείνη των ψυχολογικών παραμέτρων και συνθηκών, συγκυριών και αλληλεξαρτήσεων, ποιοτήτων και δυνατοτήτων που συνθέτουν την αφηρημένη έννοια της ανθρώπινης ικανότητας.

Ο μόνος τρόπος που έχουμε για να συνδέσουμε το κάθε ένα ερώτημα που υπάρχει μέσα σε μια ψυχομετρική κλίμακα αξιολόγησης είναι να θεωρήσουμε ότι το τελικό αποτέλεσμα αυτής της κλίμακας εκφράζει ποσοτικά ή και ποιοτικά τη συνάρτηση του λανθάνοντος χαρακτηριστικού. Για τον λόγο αυτόν, η βασική λογική της θεωρίας είναι ότι η επίδοση σε ένα ερώτημα συσχετίζεται με το λανθάνον χαρακτηριστικό του εξεταζομένου, όπως αυτό εκτιμάται (συνήθως) με βάση τη συνολική επίδοση σε όλα τα ερωτήματα της κλίμακας. Αυτή η «πρακτική» διευκολύνει με δύο τρόπους: πρώτον, μας επιτρέπει να χρησιμοποιήσουμε μια *εκτίμηση* του λανθάνοντος χαρακτηριστικού που μας ενδιαφέρει για να βρούμε τη συσχέτιση του κάθε ερωτήματος με αυτή τη συνάρτηση. Από την «πρακτική» αυτή προέρχεται και η διαδικασία του υπολογισμού των συναφειών ερωτημάτων-συνόλου. Δεύτερον, εάν τελικά αποδειχθεί ότι τα ερωτήματα της ψυχομετρικής κλίμακας αξιολόγησης λειτουργούν αξιόπιστα, έγκυρα και εκτιμούν διάφορες πλευρές της τελικής συνάρτησης του λανθάνοντος χαρακτηριστικού, ενώ ταυτόχρονα συσχετίζονται με αυτό (όπως ορίζεται από την τελική τιμή που προέρχεται από τα ερωτήματα) τότε το ψυχομετρικό μας εργαλείο μπορεί να εκτιμήσει, στο γενικότερο σύνολό του, το λανθάνον χαρακτηριστικό που μας ενδιαφέρει.

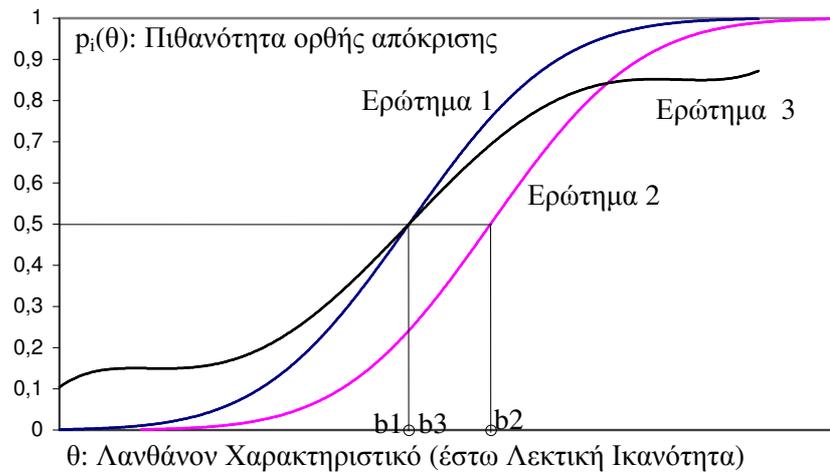
Η πιο πάνω θεώρηση, δεν αποτελεί μόνο εναλλακτική μέθοδο ελέγχου ερωτημάτων, αλλά και συνδυάζεται με τις άλλες μεθόδους ώστε τα ερωτήματα τα οποία τελικά θα χρησιμοποιηθούν στο τελικό ψυχομετρικό, προς στάθμιση, εργαλείο να είναι τα καταλληλότερα. Για να επιτύχει τον στόχο αυτόν, η θεωρία χρησιμοποιεί τις χαρακτηριστικές «καμπύλες» ερωτημάτων (*Item Characteristic Curves – ICC*). Οι

καμπύλες αυτές αναφέρονται σε κάθε ερώτημα ξεχωριστά, σε σχέση πάντα με το λανθάνον χαρακτηριστικό και «περιέχουν» 3 παραμέτρους, βασικές για την καταλληλότητα ενός ερωτήματος:

- α) την παράμετρο διαφοροποιητικής ισχύος του ερωτήματος
- β) την παράμετρο δυσκολίας του και,
- γ) την παράμετρο τύχης για ερωτήματα πολλαπλών επιλογών.

Ως προς την τρίτη παράμετρο, θα αρκεστούμε στο να επισημάνουμε ότι συνήθως εφαρμόζονται τακτικές εξάλειψης της τύχης όταν έχουν συλλεγεί τα στοιχεία για κάποιον εξεταζόμενο, αλλά για να είναι γνωστή η διαδικασία εξάλειψης της τύχης, θα πρέπει να έχει πρώτα μελετηθεί και περιγραφεί σε βάθος η αντίστοιχη παράμετρος στην χαρακτηριστική καμπύλη του κάθε ερωτήματος. Πριν προχωρήσουμε όμως στο να αναλύσουμε τις άλλες δύο θεμελιώδεις παραμέτρους, θα παρουσιάσουμε στο Σχήμα της σελίδας 91 τρεις τέτοιες χαρακτηριστικές καμπύλες για τρία ερωτήματα (έστω τρία ερωτήματα μιας ψυχομετρικής κλίμακας αξιολόγησης που επιχειρεί να μετρήσει το λανθάνον χαρακτηριστικό της λεκτικής ικανότητας).

Στο Σχήμα (σελ. 91) περιγράφεται η αθροιστική κατανομή συχνότητας (*cumulative frequency distribution*) για 3 ερωτήματα. Η μορφή των τριών αυτών χαρακτηριστικών καμπυλών είναι σιγμοειδής, δηλαδή εκφράζει μικρές σχετικές συχνότητες στα άκρα της αθροιστικής κατανομής και μεγαλύτερες στο μέσον της κατανομής. Όμως δεν έχει την ίδια ακριβώς μορφή και για τα τρία ερωτήματα ενώ επίσης, για κάθε ερώτημα, αρχίζει από διαφορετικό σημείο του οριζοντίου συνεχούς. Θα πρέπει να διευκρινίσουμε ότι ο οριζόντιος άξονας είναι το συνεχές πάνω στο οποίο αναπτύσσεται το λανθάνον χαρακτηριστικό (θ). Στην πράξη, ο άξονας αυτός παριστάνει διαφορετικά επίπεδα ενός εξωτερικού κριτηρίου (π.χ. ηλικία) που αντιπροσωπεύει το (θ). Αυτό συμβαίνει διότι δεν γνωρίζουμε αριθμητικά, μέσω μετρικής προσέγγισης, το ίδιο το λανθάνον χαρακτηριστικό. Άλλωστε, εάν το γνωρίζαμε δεν θα υπήρχε λόγος να δημιουργήσουμε την κλίμακα αξιολόγησης. Έτσι, χρησιμοποιούμε κάποιο εξωτερικό-αντικειμενικό κριτήριο το οποίο θεωρητικά είναι συναφές με το λανθάνον χαρακτηριστικό ή/και εκφράζει διάφορες διαβαθμίσεις του. Στον κάθετο άξονα παριστώνται οι p -τιμές για κάθε επίπεδο του οριζοντίου άξονα.



[Πηγή: Anastasi, A. (1990) *Psychological Testing*, έκτη έκδοση, N.Y., Macmillan Publishing Company, σελ. 222.]

Στην περίπτωση του ερωτήματος 1 παρατηρούμε ότι η σιγμοειδής χαρακτηριστική καμπύλη μοιάζει με το μισό της κανονικής κωδωνοειδούς συμμετρικής κατανομής του *Gauss*. Οι δύο άκρες της καμπύλης είναι, στην πραγματικότητα, ασυμπτωτικές, δηλαδή τείνουν να γίνουν οριζόντιες ευθείες (παράλληλες με τον οριζόντιο άξονα) αν και ποτέ δεν θα γίνουν απόλυτα οριζόντιες. Στα άκρα αυτά η διαφοροποίηση μεταξύ διαφορετικών επιπέδων (θ) ως προς την επιτυχία στο ερώτημα αυτό είναι ελάχιστη. Στο μέσον όμως της χαρακτηριστικής καμπύλης παρατηρείται σταδιακή άνοδος, άρα και υψηλή διαφοροποίηση μεταξύ των διαφορετικών επιπέδων (θ). Η ίδια χαρακτηριστική καμπύλη ισχύει και για το ερώτημα 2. Οι καμπύλες αυτές επομένως αντιπροσωπεύουν μεγάλη διαφοροποιητική ισχύ των αντίστοιχων ερωτημάτων για το μεγαλύτερο εύρος του λανθάνοντος χαρακτηριστικού. Κατά συνέπεια, μικρές αλλαγές στο λανθάνον χαρακτηριστικό θα αντικατοπτρίζονται στο βαθμό επιτυχίας για τα ερωτήματα αυτά, άρα και οι συμμετέχοντες-εξεταζόμενοι θα διαφοροποιούνται μεταξύ τους σε ικανοποιητικό βαθμό. Δεν συμβαίνει το ίδιο για το ερώτημα 3 που διαθέτει πολύ μικρότερη διαφοροποιητική ισχύ, καθώς σε όλο σχεδόν το εύρος του λανθάνοντος χαρακτηριστικού οι τιμές p παραμένουν σχεδόν ίδιες (με εξαίρεση τις μεσαίες τιμές). Το ερώτημα 3 λοιπόν δεν διαφοροποιεί μεταξύ των εξεταζομένων και ταυτόχρονα (όπως φαίνεται από το σημείο που η χαρακτηριστική καμπύλη τέμνει τον κάθετο

άξονα) υπάρχει και επίδραση τύχης (επειδή ακόμη και όταν το λανθάνον χαρακτηριστικό είναι 0 η τιμή p είναι μεγαλύτερη από το 0).

Είδαμε στο 1^ο Μέρος ότι η διαφοροποιητική ισχύς ενός ερωτήματος εξαρτάται από τη διασπορά των αντίστοιχων τιμών p . Όπως είναι προφανές και στο Σχήμα της σελ. 91, η διασπορά αυτή είναι η μέγιστη για τα ερωτήματα 1 και 2 (εξάλλου, καλύπτει όλο το εύρος του κάθετου άξονα), ενώ δεν είναι ιδιαίτερα μεγάλη για το ερώτημα 3.

Με βάση τις τιμές p και πάλι (κατά ομάδες θ), μπορούμε να υπολογίσουμε και τον δείκτη δυσκολίας του κάθε ερωτήματος σε συνάρτηση με το λαθάνον χαρακτηριστικό. Έχοντας κατασκευάσει την χαρακτηριστική καμπύλη για ένα ερώτημα, σύρουμε από το σημείο 0,50 του κάθετου άξονα (το σημείο που αντιπροσωπεύει τον θεωρητικό μέσον όρο επιτυχίας) την παράλληλη με τον οριζόντιο άξονα ευθεία, η οποία και τέμνει σε κάποιο σημείο την χαρακτηριστική καμπύλη. Από το σημείο αυτό τομής, σύρουμε παράλληλη με το κάθετο άξονα ευθεία η οποία και τέμνει τον οριζόντιο άξονα σε κάποιο σημείο. Το σημείο αυτό είναι ο δείκτης δυσκολίας του ερωτήματος *σε συνάρτηση με το λαθάνον χαρακτηριστικό και όχι μόνο σε σχέση με τα υπόλοιπα ερωτήματα*, καθώς η «ποσότητα» αυτή του λανθάνοντος χαρακτηριστικού απαιτείται για την κατά μέσο όρο ορθή απάντηση στο ερώτημα αυτό.

Παρατηρούμε ότι για το ερώτημα 1 το σημείο αυτό βρίσκεται σε σημείο που αντιπροσωπεύει μικρότερη ποσότητα του (θ) από εκείνη που αντιπροσωπεύεται από το σημείο τομής του οριζοντίου άξονα για το ερώτημα 2. Η διαπίστωση αυτή σημαίνει ότι το ερώτημα 2 είναι δυσκολότερο από το ερώτημα 1 καθώς απαιτεί ικανότερο μέγεθος (θ) για να απαντηθεί κατά μέσο όρο ορθά. Ως προς το δείκτη δυσκολίας, το ερώτημα 3 είναι επίσης το ίδιο δύσκολο με το ερώτημα 1, αλλά δεν διαθέτει την ίδια διαφοροποιητική ισχύ.

Καταλήγοντας, μέσω της κατασκευής των χαρακτηριστικών καμπυλών για όλα τα ερωτήματα, είμαστε σε θέση να τα ιεραρχήσουμε με βάση τον δείκτη δυσκολίας ως προς (θ) και να συνεκτιμήσουμε τη διαφοροποιητική ισχύ του καθενός. Όταν γίνει αυτή η διαδικασία θα έχουμε καταλήξει σε μία κλίμακα που διαθέτει

διαφοροποιητική ισχύ (θα έχουμε προφανώς επιλέξει τα ερωτήματα εκείνα που δεν υστερούν στον τομέα αυτόν) και ταυτόχρονα θα αρχίζει από το ευκολότερο ερώτημα και θα καταλήγει στο δυσκολότερο. Η ανάγκη αυτής της ιεράρχησης θα συζητηθεί περαιτέρω στο τέλος του κεφαλαίου.

Η θεωρία των λανθανόντων χαρακτηριστικών διατείνεται ότι το βασικό της επίτευγμα είναι η δυνατότητα χρήσης μιας κλίμακας αξιολόγησης που έχει προκύψει από την πιο πάνω διαδικασία σε διαφορετικού επιπέδου (θ) πληθυσμούς. Αυτό συμβαίνει, καθώς (σύμφωνα πάντα με τη θεωρία) δεν θα παρατηρείται διακύμανση στις παραμέτρους των ερωτημάτων όταν αυτά χορηγούνται σε διαφορετικού επιπέδου (θ) πληθυσμούς. Έτσι, οι κλίμακες είναι “*sample-free*” δηλαδή ανεξάρτητες από τα ειδικά χαρακτηριστικά των δειγμάτων ή των εξεταζομένων. Με άλλα λόγια, η διαφοροποιητική ισχύς των ερωτημάτων, καθώς και ο δείκτης δυσκολίας τους ως προς το λανθάνον χαρακτηριστικό δεν επηρεάζονται από τα ειδικά χαρακτηριστικά (θ) του δείγματος, ούτε και αλληλεπιδρούν με αυτά, αλλά το αντίστοιχο ερώτημα διατηρεί την ικανότητα αντικειμενικής μέτρησης των ειδικών αυτών χαρακτηριστικών (θ).

Με την ίδια λογική, επειδή τα ερωτήματα διαθέτουν διαφορετικού βαθμού δείκτες δυσκολίας, είναι δυνατόν χορηγώντας διαφορετικά τμήματα μιας κλίμακας να προσεγγίσουμε με τον ίδιο αντικειμενικό τρόπο τα λανθάνοντα χαρακτηριστικά διαφορετικών ως προς αυτά πληθυσμών. Έτσι, στο WISC-III η χορήγηση της Κλίμακας «Λεξιλόγιο» αρχίζει από την 5η ερώτηση για την ηλικία των 11 ετών. Αυτό συμβαίνει καθώς γνωρίζουμε ότι ο δείκτης δυσκολίας για τις προηγούμενες 4 ερωτήσεις για την ηλικία αυτή είναι πολύ μικρός ή μηδενικός. Δεν χρειάζεται επομένως να χορηγήσουμε από την αρχή την κλίμακα αυτή αξιολόγησης στη συγκεκριμένη ηλικία, αλλά υποθέτουμε ότι τα παιδιά αυτά απαντούν στις 4 πρώτες ερωτήσεις σωστά. Χορηγούμε λοιπόν, ουσιαστικά, ένα τμήμα της κλίμακας αξιολόγησης και προσεγγίζουμε με την ίδια αντικειμενικότητα το (θ) της ηλικίας αυτής, όπως το προσεγγίζουμε και για την ηλικία των 6 ετών, στη οποία χορηγούμε ολόκληρη την κλίμακα αξιολόγησης. Η ιεράρχηση των ερωτημάτων ανάλογα με τον δείκτη δυσκολίας ως προς (θ) μας επιτρέπει μια περαιτέρω τμηματοποίηση της κλίμακας αξιολόγησης, τη διακοπή τη χορήγησής της μετά από συγκεκριμένο αριθμό

συνεχών λαθών. Η διαδικασία αυτή αποτελεί «διευκόλυνση» και για το παιδί και για τον εξεταστή και είναι αναγκαία για πολλούς μετρικούς και ψυχολογικούς λόγους.

Τέλος, ανάμεσα στα διαφορετικά επίπεδα (θ) θα πρέπει να έχουν προβλεφθεί «γέφυρες» (*calibration items*), ρυθμιστικά δηλαδή ερωτήματα που αντιστοιχούν ως προς τον δείκτη δυσκολίας και στα δύο γειτονικά επίπεδα (θ) ώστε να καλύπτουμε και αυτό το εύρος αυτό του λανθάνοντος χαρακτηριστικού.

Το μοντέλο Rasch

Στις επόμενες παραγράφους θα παρουσιάσουμε πολύ σύντομα μια μετεξέλιξη της θεωρίας των λανθανόντων χαρακτηριστικών, η οποία βασίζεται σε πολύπλοκα μαθηματικά μοντέλα και η οποία θεωρήθηκε στην αρχή τουλάχιστον ιδιαίτερα χρήσιμη για την Ψυχομετρία, την Παιδαγωγική, την Ψυχολογία και, γενικότερα, τις Κοινωνικές επιστήμες. Το μοντέλο του θεωρητικού *Rasch* προέβλεπε ότι για ένα μεγάλο αριθμό ερωτημάτων, εφόσον αυτά συγκεντρωθούν σε μια μεγάλη βάση δεδομένων, είναι δυνατόν με επανατροφοδοτικούς μηχανισμούς να καθορίσουμε ποια από αυτά δεν επηρεάζονται από τύχη και ταυτόχρονα είναι ανεξάρτητα από το δείγμα που μας δίνει τις απαντήσεις καθώς και από όλα τα υπόλοιπα ερωτήματα της κλίμακας αξιολόγησης. Αυτή η ιδιότητα των ερωτημάτων (να είναι δηλαδή *sample-free* και *item-free*) έδωσε στο μοντέλο *Rasch* πολύ αισιόδοξη ταυτότητα, καθώς σήμαινε ότι από τη βάση των ερωτημάτων θα μπορούσαμε ανά πάσα στιγμή να επιλέξουμε ένα οποιοδήποτε σύνολο ερωτημάτων που θα διέθεταν τα πιο πάνω χαρακτηριστικά και να το χορηγήσουμε, μετρώντας αντικειμενικά, αξιόπιστα και έγκυρα το λανθάνον χαρακτηριστικό στο οποίο αναφέρονται.

Ο *Rasch* θεώρησε εξ αρχής ότι όλα τα ερωτήματα της βάσης ερωτημάτων θα διαθέτουν την ίδια διαφοροποιητική ισχύ για το συγκεκριμένο λανθάνον χαρακτηριστικό, δηλαδή θεώρησε ότι το μεγάλο αυτό σύνολο ερωτημάτων θα είναι ανεξάρτητο από την ποιότητα αυτή, καθώς θα τη διαθέτει σε όμοιο για όλα τα ερωτήματα βαθμό. Μάλιστα, υποστήριξε ότι είναι δυνατόν με διαδοχικές χορηγήσεις διαφορετικών συνόλων ερωτημάτων κάθε φορά να καταλήξουμε σε ένα δείκτη που

θα συνοδεύει το κάθε ερώτημα στη βάση ερωτημάτων και το οποίο σε κάθε χορήγηση θα αναδιαμορφώνεται ανάλογα με τα νέα ψυχομετρικά χαρακτηριστικά του ερωτήματος. Αυτός ο δείκτης θα μπορεί να μας υποδείξει τα σύνολα των προς χορήγηση ερωτημάτων ανάλογα με την περίσταση και βασίζεται κυρίως στον δείκτη δυσκολίας του ερωτήματος, ενώ αποτελεί συνυπολογιζόμενη στον τελικό βαθμό του εξεταζομένου παράμετρο. Δεν χρησιμοποιούμε δηλαδή μόνο τα ερωτήματα και βαθμολογούμε τις απαντήσεις σε αυτά, αλλά συνυπολογίζουμε και τον δείκτη που συνοδεύει το κάθε ερώτημα και έχει προκύψει από τη σταδιακή δειγματοληπτική διαδικασία.

Το μοντέλο Rasch φάνηκε να είναι ιδιαίτερα χρήσιμο, ειδικά για την Παιδαγωγική και την εκπαίδευση γενικότερα, καθώς από ένα μεγάλο σύνολο ερωτημάτων θα μπορούσαμε κάθε φορά να επιλέγουμε τα αντίστοιχα με το λανθάνον χαρακτηριστικό που μας ενδιαφέρει και να τα χορηγούμε σε να ήταν μια εξ αρχής συγκροτημένη, αυτόνομη κλίμακα αξιολόγησης.

Όμως, υπάρχουν προβλήματα και στατιστικής φύσεως αλλά και εγκυρότητας αντιπροσωπευτικού περιεχομένου. Σε πολλές περιπτώσεις, τα ερωτήματα που προκύπτουν μέσω της θεωρίας των λανθανόντων χαρακτηριστικών και χρησιμοποιούνται σε μια ψυχομετρική κλίμακα αξιολόγησης *ταυτίζονται* με εκείνα που προκύπτουν αντίστοιχα από το μοντέλο Rasch. Η ταύτιση αυτή εμπεριέχει μια στατιστική αντίφαση: πώς γίνεται τα ερωτήματα του μοντέλου Rasch που είναι ανεξάρτητα από τα υπόλοιπα ερωτήματα της κλίμακας να είναι ταυτόχρονα και εξαρτημένα από αυτά, όπως προκύπτει από την κλασσική θεωρία λανθανόντων χαρακτηριστικών;

Ο δεύτερος προβληματισμός αφορά στην εγκυρότητα αντιπροσωπευτικού περιεχομένου που θα έχει μια κλίμακα όταν έχει επιλεγεί με τη μέθοδο του Rasch. Είναι προφανές, ότι η διαδικασία υπολογισμού των χαρακτηριστικών δεικτών κάθε ερωτήματος είναι χρονοβόρα και απαιτεί πολλές δειγματοληπτικές ερευνητικές προσπάθειες. Θα περάσουν, επομένως, χρόνια εάν κάποιος προσπαθήσει να κατασκευάσει την αντίστοιχη βάση ερωτημάτων και να υπολογίσει τους χαρακτηριστικούς δείκτες. Όμως, στη διάρκεια των χρόνων αυτών αλλάζουν τα στοιχεία στα οποία αναφέρονται τα αρχικώς δημιουργηθέντα ερωτήματα. Έστω ότι

στη βάση ερωτημάτων υπάρχει ένα ερώτημα που αναφέρεται στη χρήση των ουσιαστικών στη γλώσσα. Το ερώτημα αυτό, μετά από διαδοχικές δειγματοληψίες από παιδιά τρίτης δημοτικού έχει αποκτήσει ένα ιδιαίτερα ικανοποιητικό χαρακτηριστικό δείκτη και παραμένει με την ιδιότητα αυτή στη βάση ερωτημάτων. Μετά από 5 έτη κάποιος ερευνητής θέλει να χρησιμοποιήσει τα καλύτερα δυνατά ερωτήματα για τη μέτρηση της εκφραστικής ικανότητας. Χρησιμοποιεί λοιπόν και το ερώτημα αυτό μαζί με άλλα και το χορηγεί σε παιδιά της τρίτης τάξης του δημοτικού και διαπιστώνει ότι κανείς δεν απαντά στο ερώτημα αυτό ορθά, αν και ο δείκτης δυσκολίας του είναι μικρός. Το πρόβλημα προέρχεται από το γεγονός ότι 5 έτη πριν, τα παιδιά της τρίτης δημοτικού είχαν διδαχθεί τη χρήση των ουσιαστικών στη γλώσσα, ενώ για την ερευνητική προσέγγιση που επιχειρεί ο συγκεκριμένος ερευνητής έχει αλλάξει το αναλυτικό πρόγραμμα και τα παιδιά δεν διδάσκονται πλέον το αντικείμενο. Επομένως δεν υφίσταται, στην περίπτωση, εγκυρότητα αντιπροσωπευτικού περιεχομένου της κλίμακας με όλα τα συνεπαγόμενα προβλήματα.

Πάντως, το μοντέλο Rasch, από θεωρητική άποψη έχει δώσει έναυσμα για την δημιουργία ψυχομετρικών κλιμάκων αξιολόγησης που επιχειρούν να προσεγγίσουν αυτή τη θεωρητική λογική των *sample-free & item-free* ερωτημάτων. Με τη λογική αυτή λειτουργούν οι κλίμακες μέτρησης *Guttman*. Για τις κλίμακες αυτές μέτρησης ισχύει η παραδοχή Rasch της ίσης διαφοροποιητικής ισχύος και τα ερωτήματα απλά τοποθετούνται σε σειρά μέσα στην κλίμακα αξιολόγησης με βάση τους δείκτες δυσκολίας και μόνο, ενώ επίσης, προϋποθέτουν την απουσία τύχης. Με τον τρόπο αυτό, μέσω διαδικασιών στάθμισης, είναι δυνατόν να καταλήξουμε σε κλίμακες που πληρούν τα κριτήρια της διαφοροποιητικής ισχύος για ένα εύρος λανθάνοντος χαρακτηριστικού και ταυτόχρονα προσεγγίζουν διαφορετικά επίπεδα του χαρακτηριστικού αυτού μετρώντας τα σημεία εκείνα στα οποία βρίσκεται ο κάθε εξεταζόμενος και συγκρίνοντάς τα με το όλο συνεχές του χαρακτηριστικού και τις θέσεις που κατέχουν οι άλλοι (ο γενικός πληθυσμός) στο χαρακτηριστικό αυτό.

Οι κλίμακες μέτρησης *Guttman* χρησιμοποιούνται κυρίως στη μέτρηση χαρακτηριστικών όπως ικανότητες, νοημοσύνη, γνώσεις, ταχύτητα, κ.τ.λ.. Επειδή έχουν ενσωματωμένη την παράμετρο της ορθότητας των απαντήσεων, και επειδή σε

άλλα τεστ (αξιών, στάσεων, πεποιθήσεων, κινήτρων, προσωπικότητας, ιδιοσυγκρασίας, άγχους, αυτοεκτίμησης, κ.τ.λ.) αυτή η παράμετρος δεν υφίσταται καθώς δεν υπάρχουν σωστές ή λανθασμένες απαντήσεις, οι κλίμακες αυτές και γενικότερα το μοντέλο *Guttman* δεν είναι εύκολο να χρησιμοποιηθεί.

Στις περιπτώσεις εκείνες που οι κλίμακες μέτρησης *Guttman* δεν ενδείκνυνται, θα πρέπει βεβαίως να βρεθεί εναλλακτική οδός για τη μέτρηση των λανθανόντων χαρακτηριστικών. Την εναλλακτική αυτή οδό προσφέρουν οι κλίμακες *Likert* οι οποίες δεν προϋποθέτουν τη σταδιακή μετάβαση από εύκολα σε δύσκολα ερωτήματα, θεωρώντας ότι για κάθε ερώτημα υπάρχει ένα συγκεκριμένο εύρος πιθανών απαντήσεων (4, 5, 6 ή 7 βαθμίδες, από Πάρα πολύ έως Καθόλου, από Πολύ συχνά έως Ποτέ, κ.ο.κ.). Είναι προφανές ότι οι κλίμακες *Likert* δεν είναι κλίμακες ίσων διαστημάτων, καθώς οι αποστάσεις μεταξύ των βαθμίδων δεν είναι όμοιες. Ένα άλλο πρόβλημα είναι ότι οι μετρήσεις καταλήγουν να είναι ασυνεχείς αριθμητικές (ενώ οι κλίμακες μέτρησης *Guttman* είναι δυνατό να καταλήγουν σε συνεχείς αριθμητικές μετρήσεις, π.χ. χρόνος αντίδρασης σε εύκολα έως δύσκολα ερωτήματα). Τα προβλήματα που παρουσιάζουν οι κλίμακες *Likert* αντιμετωπίζονται συνήθως με στατιστικές μεθόδους, αν και ποτέ δεν αντιμετωπίζονται απόλυτα. Καθώς όμως σε πάρα πολλές περιπτώσεις οι λειτουργικοί ορισμοί που αναφέρονται στα λανθάνοντα χαρακτηριστικά που μας ενδιαφέρουν δεν είναι δυνατό να αντιπροσωπευθούν με μετρήσεις που προκύπτουν από κλίμακες ίσων διαστημάτων ή και ίσων λόγων, οι κλίμακες *Likert* δείχνει, προς το παρόν, να είναι η μόνη εναλλακτική οδός.

Μέρος 8. Μετατροπή των αρχικών τιμών σε δευτερογενείς κλίμακες.

Το πρώτο ερώτημα που οδηγεί στην όλη διαδικασία της στάθμισης είναι το κατά πόσο είναι συγκρίσιμα τα αποτελέσματα ενός οποιουδήποτε ψυχομετρικού τεστ ως προς δύο άτομα ή δύο διαφορετικές συνθήκες ή και μεταξύ διαφορετικών ως προς τα χαρακτηριστικά πληθυσμών. Το βασικό ζήτημα που προκύπτει από τα ερωτήματα αυτά είναι το ποια είναι η πραγματική σημασία ενός «βαθμού» σε ένα τεστ για κάποιο άτομο, σε σχέση με άλλα άτομα ή με άλλους βαθμούς του ίδιου ατόμου σε

άλλα τεστ. Αν, για παράδειγμα, κάποιος έχει απαντήσει σε ένα ψυχομετρικό τεστ άγχους και καταλήξει να έχει δείκτη 28, τότε ποια είναι η πραγματική σημασία του δείκτη αυτού; Σε σχέση με τους δείκτες που αντιστοιχούν σε άλλα άτομα, π.χ. 23, 12, 43, τι σημαίνει ο βαθμός 28, δηλαδή, πού βρίσκεται το άτομο αυτό ως προς το άγχος του σε σχέση με τους άλλους; Σε σχέση με το βαθμό υπερκινητικότητας του ίδιου ατόμου, όπως προκύπτει από κάποια άλλη ψυχομετρική δοκιμασία, ποια είναι η διασύνδεση μεταξύ των δεικτών υπερκινητικότητας και άγχους και ποιος είναι ο μεγαλύτερος δείκτης; Στα ζητήματα αυτά, η περιγραφική στατιστική απαντά με τη μέθοδο της μετατροπής των αρχικών τιμών σε δευτερογενείς τιμές. Η μετατροπή αυτή γίνεται πολλές φορές για ερευνητικούς λόγους, στα πλαίσια δηλαδή της ερευνητικής στρατηγικής, όμως αποτελεί τη βάση για τη θεωρία της στάθμισης των ψυχομετρικών τιμών²².

Οι δευτερογενείς κλίμακες που χρησιμοποιούνται κατά τη μετατροπή των τιμών είναι τα *Τυπικά Πηλικά* (z-τιμές), οι *T-τιμές*, οι *Εκατοστιαίες τιμές*, οι *Stanine* τιμές, οι *IQ-τιμές*, και οι *Αναπτυξιακές Ηλικίες*. Κατά τη διαδικασία της στάθμισης θα δούμε ότι οι κλίμακες αυτές δεν είναι οι μόνες που χρησιμοποιούνται αλλά η οποιαδήποτε άλλου είδους μετατροπή σε δευτερογενείς τιμές βασίζεται συνήθως στα *Τυπικά Πηλικά*²³.

Σε όλες τις περιπτώσεις μετατροπής σε δευτερογενείς τιμές, η προσπάθεια εστιάζεται στο να τοποθετηθούν οι αρχικές μας τιμές σε αναλογική θέση ως προς τη θεωρητική κατανομή του *Gauss*. Μέσω της υπολογιστικής διαδικασίας που δεν αποτελεί αντικείμενο της τρέχουσας παρουσίασης, μπορούμε να μετατρέψουμε τις αρχικές τιμές και κατόπιν να προβούμε σε συμπεράσματα σε σχέση με τις νέες αυτές τιμές. Για την υπολογιστική διαδικασία θα πρέπει μόνο να αναφέρουμε τον υπολογιστικό τύπο μετατροπής των αρχικών τιμών σε *Τυπικά πηλικά*, καθώς θα μας είναι ιδιαίτερα χρήσιμος στο κεφάλαιο που αναφέρεται στη διαδικασία της στάθμισης. Ο απλός υπολογιστικός αυτός τύπος βασίζεται στον μέσο όρο και την

²² Για περισσότερες πληροφορίες για τις μετατροπές πρωτογενών σε δευτερογενείς τιμές, βλ. σύγγραμμα: **Ι.Ν. Παρασκευόπουλος** (1990) *Στατιστική εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμος 1, Κεφ. 7.

²³ Θα πρέπει να διευκρινίσουμε στο σημείο αυτό ότι οι μετατροπές σε δευτερογενείς τιμές διαφέρουν από τις μαθηματικές μετατροπές που γίνονται σε ερευνητικό στάδιο για την ομαλοποίηση των κατανομών συχνότητας των διαφόρων μεταβλητών (μετατροπή τετραγωνικής ρίζας, αντιστρόφων τιμών, λογαρίθμων, κ.τ.λ.).

τυπική απόκλιση της κατανομής συχνότητας από την οποία προέρχεται η αρχική τιμή, καθώς και στην ίδια την αρχική τιμή. Είναι προφανές, ότι η μετατροπή στηρίζεται, και πάλι, σε δειγματοληπτικές διαδικασίες και δεν είναι ψυχομετρικά χρήσιμη αν δεν έχουμε στοιχεία από αντιπροσωπευτικό δείγμα του πληθυσμού που μας ενδιαφέρει.

$$z = \frac{X - \bar{X}}{\hat{s}}$$

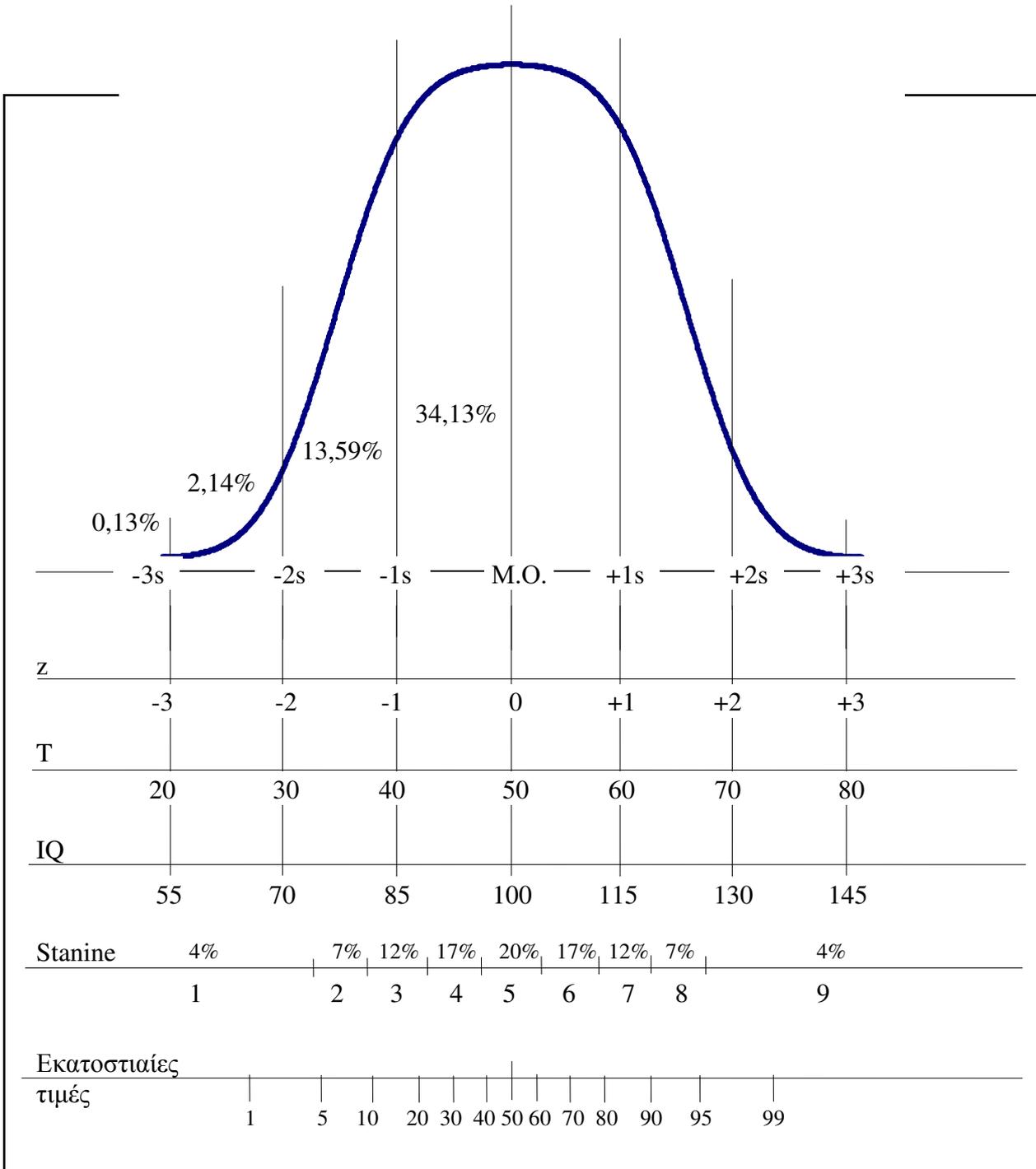
όπου X είναι η αρχική τιμή, \bar{X} είναι ο μέσος όρος της κατανομής

Κάθε δευτερογενής κλίμακα παρουσιάζει πλεονεκτήματα αλλά και μειονεκτήματα και χρησιμοποιείται ανάλογα με την περίπτωση. Συνήθως κατά τη στάθμιση ψυχομετρικών κλιμάκων νοημοσύνης ή κλιμάκων ικανοτήτων καταλήγουμε σε μετατροπή σε IQ τιμές, αναπτυξιακές ηλικίες και εκατοστιαίες τιμές, ενώ για τη στάθμιση, για παράδειγμα, ψυχομετρικών μέσων προσωπικότητας χρησιμοποιούμε συνήθως T-τιμές και εκατοστιαίες τιμές και σε άλλες περιπτώσεις χρησιμοποιούμε δευτερογενείς κλίμακες που είναι γραμμικές μετατροπές των Τυπικών πηλίκων.

Ειδική μνεία πρέπει να γίνει για τη δευτερογενή κλίμακα των αναπτυξιακών ηλικιών, η οποία είναι πολύ χρήσιμη για τα ψυχομετρικά τεστ ικανοτήτων και νοημοσύνης και η οποία προσφέρει πληροφορίες που δεν μπορούμε να συλλέξουμε από τις άλλες δευτερογενείς κλίμακες. Η βασική επιπλέον πληροφόρηση που μας παρέχει η κλίμακα των αναπτυξιακών ηλικιών είναι η αναπτυξιακή «θέση» που κατέχει ο κάθε εξεταζόμενος σε σχέση με τις αναπτυξιακές θέσεις που καταλαμβάνουν οι υπόλοιποι συμμετέχοντες στο δείγμα, και σε πλαίσια στάθμισης, ο γενικός πληθυσμός ενδιαφέροντος. Ο υπολογισμός της αναπτυξιακής καμπύλης, είναι το πρώτο βήμα για τον υπολογισμό των αναπτυξιακών ηλικιών.

Φυσικά, υπάρχει αντιστοιχία (βλ. επόμενο Σχήμα) μεταξύ δευτερογενών κλιμάκων, ενώ από το τυπικό πηλίκo, μέσω απλής γραμμικής μετατροπής, μπορούμε να εκφράσουμε την αρχική τιμή σε οποιαδήποτε δευτερογενή κλίμακα με μέσο όρο και τυπική απόκλιση της επιλογής μας. Στο WISC-III η δευτερογενής κλίμακα τυπικών βαθμών έχει μέσον όρο 10 και τυπική απόκλιση 3. Τα τυπικά πηλίκα των X πολλαπλασιάστηκαν με το 3 και στο αποτέλεσμα προσθέσαμε 10. Γενικά:

$$\text{Τυπικός βαθμός} = [(\text{Τυπικό πηλίκo αρχικής τιμής}) \otimes s] + \bar{X} \text{ δευτερογενούς κλίμακας}$$



Σημείωση: Οι ποσοστιαίες αναλογίες κάτω από την κανονική κατανομή, λόγω έλλειψης χώρου, δεν αντιστοιχούν επακριβώς στις πραγματικές, αν και η μορφή της κατανομής έχει υπολογισθεί με βάση τη μαθηματική συνάρτηση της κανονικής κωδωνοειδούς συμμετρικής κατανομής του *Karl Gauss*.

διάφορα άτομα, δεν σταματούμε στην ανίχνευση της χρονολογικής ηλικίας στην οποία αντιστοιχεί το άτομο με βάση την επίδοσή του, αλλά προχωρούμε και στον υπολογισμό του αναπτυξιακού του πηλίκου, ώστε να γνωρίζουμε και τη μελλοντική εξέλιξη του ατόμου στο λανθάνον χαρακτηριστικό που μας ενδιαφέρει, έστω στη μηχανική ικανότητα.

Η Ψυχομετρία δεν σταματά εδώ, ούτε και μπορεί κανείς να ισχυρισθεί ότι γνωρίζοντας τα όσα έχουν γίνει αντικείμενο διαπραγμάτευσης μέχρι εδώ έχει και την ικανότητα να δημιουργήσει, να ελέγξει και να χρησιμοποιήσει ένα μέσο συλλογής δεδομένων, ψυχομετρικό εργαλείο, ή «τεστ».

Μπορεί, όμως, να δοκιμάσει.

Ενδεικτική ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Γεώργας, Δ.Δ. (1971). *Georgas Τεστ Νοημοσύνης για Παιδιά*, Αθήνα, Κέδρος.

Γεώργας, Δ.Δ., Παρασκευόπουλος, Ι.Ν., Μπεζεβέγκης, Η.Γ., & Γιαννίτσας, Ν.Δ. (1997). *Ελληνικό WISC-III: Wechsler Κλίμακες Νοημοσύνης για Παιδιά, Οδηγός Εξεταστή*. Πανεπιστήμιο Αθηνών, Φιλοσοφική Σχολή, Ψυχομετρικό Εργαστήριο. Αθήνα, Ελληνικά Γράμματα.

Παρασκευόπουλος, Ι.Ν. (1990). *Στατιστική εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμοι 1 και 2. Αθήνα.

Παρασκευόπουλος, Ι.Ν. (1992). *Ψυχολογία Ατομικών Διαφορών*. Αθήνα.

Παρασκευόπουλος, Ι.Ν. (1993). *Μεθοδολογία Επιστημονικής Έρευνας*, Τόμοι 1 και 2. Αθήνα.

Παρασκευόπουλος, Ι.Ν., Καλαντζή-Αζίζι, Α., & Γιαννίτσας, Ν.Δ. (1999). *Το ΑθηνάΤεστ Διάγνωσης Δυσκολιών Μάθησης: Αναθεωρημένη και Επαυξημένη Έκδοση*. Αθήνα: Ελληνικά Γράμματα

Ξενόγλωσση

Anastasi, A. (1990). *Psychological Testing*, έκτη έκδοση, New York, Macmillan.

Christensen, L.B. (1988). *Experimental Methodology*, τέταρτη έκδοση. Boston: Allyn & Bacon

Hambleton, R.K., & Slater, S.C. (1997). Item Response Theory Models and Testing Practices: Current international status and future directions, *European Journal of Psychological Assessment*, 13 (1), 21-28.

Kline, P. (1993). *The Handbook of Psychological Testing*. London: Routledge.

Rust, J., & Golombok, S. (1989). *Modern psychometrics: The science of Psychological Assessment*. London: Routledge.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using Multivariate Statistics*, δεύτερη έκδοση. New York: Harper & Row.