

Μεθοδολογία στάθμισης μιας κλίμακας

Β. Ραφτόπουλος¹,
Θ. Θεοδοσοπούλου²

¹ΤΕΙ Λάρισας, ²Τμήμα Νοσηλευτικής,
Πανεπιστήμιο Αθηνών, Αθήνα

Scale validation methodology

Abstract at the end of the article

1. Εισαγωγή
2. Εγκυρότητα (Validity)
 - 2.1. Προβλεπτική εγκυρότητα
 - 2.2. Συντρέχουσα εγκυρότητα
 - 2.3. Εγκυρότητα περιεχομένου
 - 2.4. Εγκυρότητα εννοιολογικής κατασκευής
 - 2.5. Η ποιοτική έρευνα ως προαπαιτούμενο για την εγκυρότητα εννοιολογικής κατασκευής
3. Αξιοπιστία (Reliability)
 - 3.1. Τα είδη της αξιοπιστίας μιας κλίμακας
4. Παραγοντική ανάλυση (Factor Analysis)
 - 4.1. Είδη της παραγοντικής ανάλυσης
 - 4.2. Μορφές της παραγοντικής ανάλυσης
 - 4.3. Κριτήρια για τον προσδιορισμό των παραγόντων
 - 4.4. Επάρκεια της δειγματοληψίας
 - 4.5. Στρέψη των αξόνων

Λέξεις ευρετηρίου

Αξιοπιστία
Εγκυρότητα
Ικανοποίηση ασθενών
Παραγοντική ανάλυση
Στάθμιση κλίμακας

Υποβλήθηκε 27.11.2001

Εγκρίθηκε 6.3.2002

1. ΕΙΣΑΓΩΓΗ

Για τις κοινωνικές επιστήμες, η περιγραφή της ανθρώπινης συμπεριφοράς και των εκφάνσεών της αποτέλεσε ένα διακαή αυτοσκοπό. Στην πορεία των ετών, το απόλυτο εργαλείο για την επίτευξη αυτού του στόχου ήταν η δοκιμασία (test), δηλαδή η συστηματική διαδικασία παρατήρησης και περιγραφής της ανθρώπινης συμπεριφοράς, με τη βοήθεια μιας αριθμητικής κλίμακας.¹ Η ερμηνευτική λειτουργικότητα της δοκιμασίας διαφαίνεται μέσα από τη διαδικασία της στάθμισής της, δηλαδή της διασφάλισης της ομοιομορφίας της και της επαναληψιμότητάς της με τον ίδιο περιγραφικό τρόπο, σε διαφορετικό τόπο, χρόνο και ομάδα του πληθυσμού. Αυτά τα χαρακτηριστικά μπορούν να επιβεβαιωθούν με την εκτίμηση των ψυχομετρικών ιδιοτήτων της δοκιμασίας, δηλαδή της εγκυρότητας και της αξιοπιστίας της.

2. ΕΓΚΥΡΟΤΗΤΑ (VALIDITY)

Κάθε ψυχομετρική κλίμακα πρέπει να διακρίνεται για δύο βασικές ιδιότητές της: την αξιοπιστία (reliability) και την εγκυρότητά της. Ο Hammersley υποστηρίζει ότι: «ένα εργαλείο είναι έγκυρο ή αληθές, εάν αντιπροσωπεύει

επακριβώς τις διαστάσεις ενός φαινομένου, που σκοπεύει να περιγράψει, να εξηγήσει ή να διαμορφώσει σε θεωρητικό επίπεδο».² Στις περισσότερες αναλύσεις, η εγκυρότητα σχετίζεται με την αξιοπιστία.^{2,3} Η εγκυρότητα συνίσταται στο κατά πόσο η κλίμακα μετρά αυτό που επικαλείται ότι μετρά. Κατά τον Cronbach, το βασικό κριτήριο για την επιλογή μιας κλίμακας είναι η εγκυρότητά της.¹ Για παράδειγμα, αν μετράμε την ικανοποίηση του ασθενούς από την ποσότητα της πληροφόρησης που λαμβάνει για την εξέλιξη της υγείας του, πρέπει να βεβαιωθούμε ότι δεν μετράμε την ικανοποίηση του ασθενούς από τη στάση του νοσηλευτή αναφορικά με την πληροφόρηση του ασθενούς.

Στην παραδοσιακή φιλοσοφία, η εγκυρότητα αναφέρεται στην *ορθότητα* των απαντήσεων παρά στην *αλήθεια* τους. Για παράδειγμα, η δήλωση «όλα τα άλογα έχουν δύο κεφάλια, το X είναι ένα άλογο, επομένως το X έχει δύο κεφάλια» είναι μια έγκυρη δήλωση, αλλά όχι και αληθής. Η αναλήθεια της πρώτης πρότασης «όλα τα άλογα έχουν δύο κεφάλια» προσδιορίζεται μέσα από την ικανότητά της να κερδίσει τη συγκατάβαση μιας κοινότητας (ομάδας), η οποία φαίνεται να ακολουθεί μια χρονικά εξελικτική διαδικασία, δεδομένου ότι τα φαινόμενα εναλλάσσονται μεταξύ αλήθειας και μη στο διάβα των

χρόνων. Από την άλλη μεριά, επειδή η αλήθεια μιας δήλωσης βασίζεται στην αντίληψή της, η εγκυρότητα της δεν οριοθετείται κατ' ανάγκη από τη λογική ερμηνεία της δήλωσης. Έτσι, το αν το άλογο έχει δύο κεφάλια εξαρτάται από το πώς ένας άνθρωπος με ένα συγκεκριμένο κοινωνικοπολιτισμικό και ιστορικό προφίλ αντιλαμβάνεται τους όρους «κεφάλι» και «άλογο». Στην ουσία, μάλλον πρόκειται για έναν «κανονιστικό εννοιολογικό συμβιβασμό» στις καθημερινές απαιτήσεις της κάθε εποχής, με κριτήριο τη διασφάλιση της επικοινωνίας των ανθρώπων. Για το λόγο αυτό, χρησιμοποιούμε κάποια κριτήρια –δηλαδή ορίζουμε μια κατεύθυνση– για την αξιολόγηση της εγκυρότητας των δηλώσεων. Εάν λοιπόν μια δήλωση γίνεται με τρόπο ώστε να ανταποκρίνεται στα συγκεκριμένα κριτήρια εγκυρότητας, τότε αντιμετωπίζεται ως αληθής, για το χρονικό διάστημα που κερδίζει τη συμφωνία της κοινότητας με το συγκεκριμένο προφίλ.

Η εγκυρότητα των ψυχομετρικών κλιμάκων εμφανίστηκε ως όρος μετά την πρωτοβουλία της Επιτροπής του Αμερικανικού Συνδέσμου Ψυχολόγων APA (1950–1954) να προσδιορίσει ποια χαρακτηριστικά θα έπρεπε να αξιολογηθούν πριν δημοσιευθεί μια κλίμακα.⁴ Έτσι, προέβη σε συστάσεις, προσδιορίζοντας 4 διαφορετικούς τύπους εγκυρότητας, με προεξάρχουσα την *εγκυρότητα δομής*⁵ (construct validity), που την εμπνεύστηκε μια υποεπιτροπή του Συνδέσμου. Αυτοί οι 4 τύποι εγκυρότητας είναι:

- Η *προβλεπτική ή προγνωστική εγκυρότητα* (predictive validity)
- Η *συντρέχουσα εγκυρότητα* (concurrent validity)
- Η *εγκυρότητα περιεχομένου* (content validity)
- Η *εγκυρότητα εννοιολογικής κατασκευής* (construct validity).

Οι πρώτοι δύο τύποι εγκυρότητας θεωρούμε ότι προσανατολίζονται προς ένα κριτήριο^{6,7} (criterion-related validity). Αναφερόμενος σε αυτή την κατηγορία εγκυρότητας, ο Campbell χαρακτηρίζει την προβλεπτική και τη συγχρονική εγκυρότητα ως «πρακτική εγκυρότητα»⁸ (practical validity).

2.1. Προβλεπτική εγκυρότητα

Μια ψυχομετρική κλίμακα διακρίνεται για την προβλεπτική της εγκυρότητα στο βαθμό που είναι σε θέση, με βάση τις σημερινές αξιολογήσεις, να προβλέψει μελλοντικές παραμέτρους, που αποτελούν ειδοποιά χαρακτηριστικά της σημερινής εννοιολογικής κατασκευής. Για παράδειγμα, το κριτήριο για την εκτίμηση της προβλεπτι-

κής εγκυρότητας μιας κλίμακας που εκτιμά την ικανοποίηση των ασθενών από την ποιότητα της παρεχόμενης φροντίδας είναι η δυνατότητα πρόβλεψης της μελλοντικής συμπεριφοράς του ασθενούς, με την ερώτηση «θα προτείνατε αυτή την υπηρεσία σε ένα φίλο σας;», δεδομένου ότι η τάση του ανθρώπου προς την εκδήλωση μιας συμπεριφοράς και η μελλοντική του συμπεριφορά θεωρούνται συνώνυμοι όροι. Η Smith σημειώνει ότι, επειδή αυτή η τελευταία συσχέτιση σπάνια επιβεβαιώνεται, είναι πιο ακριβές να θεωρούμε αυτή την ερώτηση περισσότερο ως ένδειξη συγχρονικής εγκυρότητας και συμπληρώνει ότι ένας παράγοντας που καθιστά προβληματική την προβλεπτική ικανότητα των κλιμάκων για την ικανοποίηση των ασθενών είναι η απουσία μιας διακύμανσης στις απαντήσεις που προσπαθούν να αποτυπώσουν τη μελλοντική συμπεριφορά των ασθενών.⁹ Έτσι, η μεγάλη πλειοψηφία των ασθενών τείνει να προτιμά τις ακραίες απαντήσεις, δηλαδή να δηλώνουν πολύ ικανοποιημένοι ή πολύ δυσαρεστημένοι με τις περισσότερες διαστάσεις της παρεχόμενης φροντίδας.

2.2. Συντρέχουσα εγκυρότητα

Μια ψυχομετρική κλίμακα διακρίνεται για τη συντρέχουσα εγκυρότητά της, όταν οι τιμές της (score) διαφοροποιούνται προς την αναμενόμενη κατεύθυνση μεταξύ ομάδων ατόμων, οι οποίες αντικειμενικά διαφέρουν ως προς την αξιολογούμενη εννοιολογική κατασκευή. Ο έλεγχος της συντρέχουσας εγκυρότητας γίνεται στη βάση της συνάφειας της αξιολογούμενης κλίμακας με ένα αντικειμενικό κριτήριο της εννοιολογικής κατασκευής για την οποία τη χρησιμοποιούμε. Βέβαια, ο έλεγχος μπορεί να γίνει και στη βάση της συνάφειας της αξιολογούμενης κλίμακας με μια άλλη σταθμισμένη κλίμακα τεκμηριωμένης αξιοπιστίας και εγκυρότητας, που αποτυπώνει την ίδια εννοιολογική κατασκευή με την αξιολογούμενη κλίμακα. Έτσι, ο συντελεστής συσχέτισης εκλαμβάνεται ως δείκτης της συγκλίνουσας εγκυρότητας της αξιολογούμενης κλίμακας. Η μέθοδος ROC-ανάλυση χρησιμοποιείται για την εκτίμηση της συντρέχουσας εγκυρότητας. Για παράδειγμα, αν ένα νοσοκομείο ενδιαφέρεται να ταξινομήσει τους ασθενείς του σε αυτούς που θα το επισκεφτούν ξανά ή όχι στο μέλλον, η καμπύλη ROC δίνει πληροφορίες για την αξιολόγηση των μεθόδων που χρησιμοποιήσε για να προβλέψει τη μελλοντική τους συμπεριφορά.

2.3. Εγκυρότητα περιεχομένου

Ένα ψυχομετρικό μέσο διακρίνεται για την εγκυρότητα του περιεχομένου του, όταν απαντά στην ερώτηση

«οι ερωτήσεις του μέσου καλύπτουν το εύρος του περιεχομένου της εννοιολογικής κατασκευής που υποστηρίζουν;». Αυτό επιτυγχάνεται στα αρχικά στάδια σχεδιασμού της έρευνας, με την ανασκόπηση της σχετικής βιβλιογραφίας, την ποιοτική έρευνα των ατόμων που αφορά και την ανάλυση των απόψεων των ειδικών στο θέμα.

Μια άλλη διάσταση της εγκυρότητας του περιεχομένου του μέσου είναι η εμπειρική εγκυρότητα (empirical validity), σύμφωνα με την οποία εξετάζεται αν και κατά πόσο οι υποκλίμακες (subscales) της συνολικής κλίμακας σχετίζονται με άλλες μεταβλητές, που αποτελούν βασικές διαστάσεις της εννοιολογικής κατασκευής με τις οποίες θα έπρεπε να σχετίζονται. Για παράδειγμα, οι Meterko et al χρησιμοποίησαν ως δείκτες εγκυρότητας της κλίμακάς τους, για την ικανοποίηση των ασθενών, τις επιμέρους κλίμακες που εκτιμούν τη συνολική εκληφθείσα ποιότητα της φροντίδας και την πρόθεση των ασθενών να επιστρέψουν στο νοσοκομείο, αλλά και να το προτείνουν σε κάποιο γνωστό τους.¹⁰ Αναζήτησαν το βαθμό συσχέτισης αυτών των δεικτών εγκυρότητας με τις επιμέρους κλίμακες της αξιολογούμενης κλίμακας.

2.4. Εγκυρότητα εννοιολογικής κατασκευής

Η εγκυρότητα της εννοιολογικής κατασκευής μιας κλίμακας αντικατοπτρίζει το πόσο καλά κάποιος αποτύπωσε τις ιδέες του ή τη θεωρία του σε μια κλίμακα.¹¹ Για να είμαστε περισσότερο αναλυτικοί, ας αναφέρουμε ένα παράδειγμα. Όλοι μας έχουμε ακούσει στο δικαστήριο την εντολή του δικαστή «πείτε την αλήθεια, την πάσα αλήθεια και μόνο την αλήθεια». Άραγε, για ποιο λόγο δεν αρκεί η φράση «πείτε μας την αλήθεια»; Γιατί κάτι τέτοιο δεν εγγυάται ότι ο μάρτυρας θα μας πει οτιδήποτε γνωρίζει ότι είναι αλήθεια. Ίσως δεν μας πει κάποια σημαντικά πράγματα, χωρίς αυτό να σημαίνει ότι δεν μας λέει την αλήθεια. Αν λοιπόν παραφράσουμε τον όρκο του δικαστηρίου, θα έχουμε «την εννοιολογική κατασκευή, όλη την κατασκευή και μόνο την κατασκευή». Τι σημαίνει αυτό; Ας υποθέσουμε ότι έχουμε πέντε διαστάσεις μιας εννοιολογικής κατασκευής που ονομάζεται «ικανοποίηση του ασθενούς». Οι περισσότεροι θα έλεγαν ότι αυτές οι διαστάσεις είναι παρόμοιες. Αν θέλαμε, λοιπόν, να αναπτύξουμε μια κλίμακα εκτίμησης της ικανοποίησης, θα λέγαμε «την ικανοποίηση, όλη την ικανοποίηση και μόνο την ικανοποίηση». Επειδή όμως η ικανοποίηση ως έννοια αλληλοκαλύπτεται με άλλες, πιο λειτουργικό είναι να δημιουργήσουμε εξ αρχής ένα «δίκτυο εννοιών» (net of meaning), που σημαίνει ότι θα πρέπει να προσδιορίσουμε με ποια «έννοια» η κατασκευή μας συγκλίνει ή αποκλίνει. Έτσι, πρέπει να αποδείξει

κανείς ότι τα δεδομένα υποστηρίζουν τη θεωρία που ερμηνεύει την «έννοια».

Η εγκυρότητα της εννοιολογικής κατασκευής περιλαμβάνει δύο κατηγορίες:

- Εγκυρότητα μετάφρασης
 - Εγκυρότητα προσώπου (face validity)
 - Εγκυρότητα περιεχομένου (content validity)
- Εγκυρότητα με βάση το κριτήριο
 - Παραγοντική εγκυρότητα (factorial validity)
 - Συγκλίνουσα εγκυρότητα (convergent validity)
 - Διακρίνουσα εγκυρότητα (discriminant validity)
 - Εγκυρότητα με βάση την παραδοχή (assumed validity).

2.4.1. Συγκλίνουσα και διακρίνουσα εγκυρότητα. Συγκλίνουσα εγκυρότητα είναι ο βαθμός στον οποίο η αξιολογούμενη κλίμακα σχετίζεται (συγκλίνει) με άλλες κλίμακες που εκτιμούν το ίδιο φαινόμενο, την ίδια εννοιολογική κατασκευή. Ο έλεγχος της συγκλίνουσας και της διακρίνουσας εγκυρότητας μιας πολυδιάστατης ψυχομετρικής κλίμακας γίνεται με τη μέθοδο των Campbell και Fiske¹² (πίνακας πολυχαρακτηριστικών-πολυμέσων, multitrait-multimethod matrix), όπως αυτή τροποποιήθηκε στη συνέχεια από τους Marsh και Smith.¹³

Τα συνήθη κριτήρια για την εκτίμηση της συγκλίνουσας εγκυρότητας είναι:

- Ο βαθμός της συσχέτισης της κάθε ερώτησης της κλίμακας με τον αντίστοιχο παράγοντα που ερμηνεύει την εννοιολογική κατασκευή (σε όρους εσωτερικής συνέπειας).
- Οι ερωτήσεις περιέχουν το ίδιο ποσοστό πληροφορίας αναφορικά με την αξιολογούμενη κατασκευή.
- Οι ερωτήσεις είναι ισχυρότεροι δείκτες της υποτιθέμενης εννοιολογικής κατασκευής, σε σχέση με άλλες κατασκευές (έλεγχος της διακρίνουσας εγκυρότητας των ερωτήσεων της κλίμακας).
- Ο Churchill υποστηρίζει ότι η συγκλίνουσα και η διακρίνουσα εγκυρότητα θα πρέπει να εκτιμώνται ταυτόχρονα με την εκτίμηση της εγκυρότητας της εννοιολογικής κατασκευής.⁶

2.4.2. Εγκυρότητα με βάση την παραδοχή. Αυτό το είδος της εγκυρότητας εξετάζει κατά πόσο η εννοιολογική κατασκευή, την οποία προσδιορίζει η κλίμακα που αξιολογούμε, σχετίζεται με άλλες εννοιολογικές κατασκευές κατά τρόπο σύμφωνο με μια παραδοχή.¹⁴

2.5. Η ποιοτική έρευνα ως προαπαιτούμενο για την εγκυρότητα εννοιολογικής κατασκευής

Στο σημείο αυτό είναι πολύ σημαντικό να τονίσουμε τον καταλυτικό ρόλο της ποιοτικής έρευνας (qualitative research) στη δόμηση της εννοιολογικής κατασκευής και του περιεχομένου μιας κλίμακας που την αποτυπώνει. Στις μέρες μας, το να προσπαθεί κανείς να μάθει κάποια σημαντικά πράγματα για την κοινωνική έρευνα *μοιάζει σαν να περπατά σε ένα δωμάτιο με πολλούς ανθρώπους που θορυβούν. Ένα δωμάτιο γεμάτο από μικρές ομάδες, καθεμιά από τις οποίες χρησιμοποιεί τη δική της ακατάλληπη γλώσσα και το δικό της τρόπο έκφρασης*. Οι έρευνες που χρησιμοποίησαν ποιοτική έρευνα διαπίστωσαν ότι οι ασθενείς βιώνουν μια δυσκολία να ονομάσουν αυτό που εκλαμβάνουν ως ικανοποίηση ή ως δυσαρέσκεια. Έτσι, συνηθίζουν να μιλούν για κάποιες προβληματικές πτυχές της φροντίδας, αλλά η διεθνής μαρτυρία υποστηρίζει ότι τελικά αποφεύγουν να προβούν σε αξιολογικές κρίσεις για το προσωπικό. Σε μια μετα-ανάλυση των ερευνών για την ικανοποίηση των ασθενών, που υπάρχουν στη διεθνή βιβλιογραφία, το συμπέρασμα είναι ότι οι έρευνες αυτές έχουν χαμηλή αξιοπιστία και εγκυρότητα.¹⁵ Αυτή βέβαια η διαπίστωση αφορά τις έρευνες για τις οποίες υπάρχει αναφορά στις ψυχομετρικές ιδιότητες της κλίμακας που χρησιμοποιούν και, σύμφωνα με τη μετα-ανάλυση, κάτι τέτοιο αφορά μόνο το 40% των ερευνών. Το απογοητευτικό σε όλη αυτή την υπόθεση είναι ότι οι ασθενείς, ως χρήστες των υπηρεσιών φροντίδας υγείας, ενεπλάκησαν στη διαμόρφωση των ερωτήσεων των σχετικών ερωτηματολογίων μόνο στο 11% των ποσοτικών ερευνών, ενώ οι συγγενείς των ασθενών στο 3%.

Σε μια ευρεία ανασκόπηση 195 ερευνών που δημοσιεύθηκαν το 1994, στο 93% των ερευνών τα αποτελέσματα ήταν πρωτίστως ποσοτικά, ενώ στο 7% ήταν πρωτίστως ποιοτικά.¹⁶ Αυτή η μικρή αναλογία των ερευνών που χρησιμοποιούν πρωταρχικά ποιοτική έρευνα ήταν απροσδόκητη και άκρως απογοητευτική, καθώς η ποιοτική έρευνα στοχεύει στην ανάπτυξη μιας έγκυρης και αξιόπιστης θεωρίας με επαγωγικό τρόπο, που καθοδηγεί και θεμελιώνεται μέσω της έρευνας. Το αδύναμο σημείο της αξιολογικής προσέγγισης αυτών των ερευνών έγκειται σε αυτό που έγραψε ο van Maanen το 1984, ότι «η ποιότητα είναι μια αφηρημένη έννοια που αμφιταλαντεύεται ανάμεσα στο επιθυμητό και στην πραγματικότητα».¹⁷

Οι μέθοδοι της ποιοτικής έρευνας είναι γνωστές στις κοινωνικές επιστήμες και λιγότερο στις επιστήμες φροντίδας υγείας, όπου συνήθως χρησιμοποιείται η πειραματική και η ποσοτική έρευνα. Αυτή ακριβώς η προτίμη-

ση της κλινικής, βιολογικής και επιδημιολογικής έρευνας έχει ως αποτέλεσμα να θεωρείται η ποιοτική έρευνα χωρίς «επιστημονικό» έρεισμα, καθώς πιστεύεται ότι τα ποιοτικά δεδομένα είναι οπωσδήποτε υποκειμενικά και άρα εμπεριέχουν σφάλμα. Οι Kirk και Miller ορίζουν την ποιοτική έρευνα ως «μια παράδοση στην κοινωνική επιστήμη, που εξαρτάται από το να βλέπεις τους ανθρώπους στο φυσικό τους περιβάλλον, χρησιμοποιώντας τη δική τους γλώσσα και τους δικούς τους όρους».¹⁸ Γι' αυτό και η ποιοτική έρευνα αποκαλείται *νατουραλιστική*.

Τις περισσότερες φορές, οι όροι «ποιοτική έρευνα» και «ποιοτικές μέθοδοι έρευνας» χρησιμοποιούνται αδιάκριτα. Συνήθως η ποιοτική έρευνα ορίζεται αναφορικά με την ποσοτική έρευνα, με τη διαφορά ότι η ποιοτική έρευνα δεν «μετρά». Ωστόσο, παρόλο που η ποιοτική έρευνα έχει να κάνει με λέξεις και προτάσεις παρά με αριθμούς, αυτό δεν σημαίνει ότι είναι χωρίς μέτρηση ή ότι δεν μπορεί να χρησιμοποιηθεί για να εξηγήσει τα «κοινωνικά» φαινόμενα. Έτσι, η *μέτρηση* στην ποιοτική έρευνα συνήθως αναφέρεται στην ταξινόμηση ή την ταξινομία των νοημάτων. Η ποιοτική έρευνα απαντά σε ερωτήσεις όπως «τι είναι το Χ και πώς το Χ μεταβάλλεται σε διαφορετικές καταστάσεις και γιατί;», παρά σε ερωτήσεις του τύπου «πόσο μεγάλο είναι το Χ και πόσα Χ υπάρχουν;». Ουσιαστικά, η ποιοτική έρευνα αναδεικνύει τη σημασία που αποδίδει κάθε άτομο στις εμπειρίες του από την κοινωνική διάδραση και τον τρόπο που ερμηνεύει το κόσμο. Γι' αυτόν ακριβώς το λόγο καλείται και *ερμηνευτική* έρευνα. Η ποιοτική έρευνα δεν είναι χρήσιμη μόνο στο πρώτο στάδιο της ποσοτικής έρευνας. Έχει επίσης να διαδραματίσει σπουδαίο ρόλο και στην εγκυρότητά της ή στην απόδοση μιας διαφορετικής ερμηνείας του ίδιου φαινομένου, αφού πολλές φορές ενδυναμώνει την επανερμηνεία των αποτελεσμάτων της ποσοτικής έρευνας.

Η ποιοτική έρευνα είναι μια επαγωγική διαδικασία, η οποία χρησιμοποιεί την παρατήρηση και τη συνέντευξη ως εργαλεία για τη διατύπωση της ερευνητικής υπόθεσης και όχι για τον έλεγχό της. Οι Chenitz και Swanson τονίζουν πως οι όροι «εγκυρότητα» και «αξιοπιστία» αποφεύγονται στην ποιοτική έρευνα και προτιμώνται οι όροι «αλήθεια» (truth) και «ακρίβεια»¹⁹ (accuracy). Έτσι, η μεθοδολογία της ποιοτικής έρευνας στοχεύει στην κωδικοποίηση και κατηγοριοποίηση των δεδομένων και στην αναζήτηση της σχέσης τους με το εννοιολογικό υπόδειγμα (μοντέλο) του ερευνητή. Ο ερευνητής, προτού «μετρήσει» τον αριθμό των ικανοποιημένων ασθενών, πρέπει να ρωτήσει «τι είναι ποιότητα της φροντίδας και τι ικανοποίηση;» τους εμπλεκόμενους στη διαδικασία της παροχής φροντίδας υγείας. Μόνο έτσι θα μπορέσει να

διακρίνει πώς διαφοροποιείται ο ορισμός της ικανοποίησης και της ποιότητας της φροντίδας μεταξύ των διαφορετικών ανθρώπων.

Οι Fiske και Taylor υποστηρίζουν ότι οι ερευνητές συνήθως πέφτουν θύματα του «σφάλματος της λανθασμένης αντίληψης», καθώς πιστεύουν ότι αυτό που αντιλαμβάνονται και αισθάνονται συμβαδίζει με αυτό που οι ερευνώμενοι αντιλαμβάνονται και αισθάνονται.²⁰ Ο Strasser τονίζει ότι αυτό το σφάλμα αποτελεί έναν πραγματικό κίνδυνο για το σχεδιασμό της έρευνας, καθώς ο σχεδιαστής της ενδέχεται να χάσει τη «φαινομενολογία» των γεγονότων που συμβαίνουν κατά τη διάρκεια της φροντίδας και ερμηνεύονται από τον ίδιο τον ασθενή.²¹ Συνεπώς, μια έρευνα είναι *πραγματικά* έγκυρη, μόνο εάν συλλαμβάνει και αποτυπώνει τις πτυχές, τα «φαινόμενα» της φροντίδας που προσπαθούμε να αξιολογήσουμε. Αυτά τα φαινόμενα θα πρέπει να διατυπώνονται όπως ακριβώς τα βιώνει ο ασθενής και όχι όπως πιστεύει ο ερευνητής, διότι σε αυτή την περίπτωση η έρευνα δεν είναι έγκυρη από τη μεριά του ασθενούς, οπότε αυτοκαταλύεται η χρηστική σκοπιμότητά της. Για παράδειγμα, μια απάντηση στην ερώτηση «πόσο ικανοποιημένος είστε από την ενεργό συμμετοχή σας στη φροντίδα που σας παρέχεται;» ενδέχεται να μην αντανάκλα την πραγματικότητα, αφού κανείς δεν έχει ρωτήσει τους ασθενείς εάν πράγματι θέλουν να συμμετάσχουν στη λήψη αποφάσεων που αφορούν την εξέλιξη της φροντίδας τους και αν τελικά αυτή η βούληση συναρτάται με μια ποιοτική φροντίδα. Για τον ερευνητή, στη φράση «είστε ικανοποιημένος με την ανταπόκριση του προσωπικού στο κάλεσμά σας;» ο όρος ανταπόκριση για τον ασθενή μπορεί να σημαίνει «βρίσκεται κοντά μου αμέσως» ή «έρχεται όταν ζητώ τη βοήθειά του» ή «απαντά σε όλες τις ερωτήσεις μου χωρίς δισταγμό».

3. ΑΞΙΟΠΙΣΤΙΑ (RELIABILITY)

Ο Rowley τονίζει «ίσως έχει μεγαλύτερη σημασία να επιβεβαιώσουμε ότι μια δοκιμασία δεν είναι ούτε αξιόπιστη ούτε αναξιόπιστη. Μια δοκιμασία μπορεί να μας δίνει κάποιες τιμές που είναι αξιόπιστες και άλλες τιμές που είναι αναξιόπιστες»²². Την ίδια στιγμή, οι Crocker και Algina υποστηρίζουν: «μια τιμή δεν είναι αξιόπιστη ή αναξιόπιστη. Αντίθετα, η αξιοπιστία ή αναξιόπιστη είναι μια ιδιότητα των τιμών στη δοκιμασία, για μια συγκεκριμένη ομάδα εξεταζομένων».²³ Οι Gronlund και Linn διατείνονται «η αξιοπιστία αναφέρεται στα αποτελέσματα που επιτυγχάνει ένα εργαλείο αξιολόγησης και όχι στο ίδιο το εργαλείο. Συνεπώς, είναι περισσότερο λειτουργι-

κό να μιλάμε για αξιοπιστία των τιμών της δοκιμασίας ή της μέτρησης παρά της δοκιμασίας ή του εργαλείου».²⁴ Όλες αυτές οι σκέψεις οδηγούν στο συμπέρασμα ότι οι ίδιοι οι ερωτώμενοι επηρεάζουν την αξιοπιστία των τιμών και άρα είναι οξύμωρο να αναφερόμαστε στην «αξιοπιστία της δοκιμασίας» χωρίς να επικεντρωνόμαστε σε ποιους δόθηκε ή σε άλλες πτυχές του πρωτοκόλλου. Αυτές οι επιστημονικές οδήγησαν στην ανάπτυξη της «γενίκευσης της αξιοπιστίας» (reliability generalization) από τη Vacha-Haase.²⁵ Συνεπώς, μια κλίμακα μπορεί να δίνει καλά αποτελέσματα σε ένα χώρο μελέτης και ανεπαρκή σε άλλον.

Ο συντελεστής εσωτερικής συνάφειας Cronbach's alpha, που αναφέρεται κατά κόρον στη βιβλιογραφία των ερευνών και ιδιαίτερα σε ό,τι αφορά τη μελέτη της προσωπικότητας, θεωρείται ένας δείκτης αξιοπιστίας.²⁶⁻²⁸ Ο Cortina ισχυρίζεται ότι δεν υπάρχει μια διεθνής συμφωνία για την ερμηνεία του Cronbach's alpha, καθώς και για το αποδεκτό εύρος των τιμών του.

Μέσα από τη βιβλιογραφία μπορεί κανείς να επισημάσει τα εξής:

- Διαφαίνεται η θετική συσχέτιση του αριθμού των ερωτήσεων μιας κλίμακας και της τιμής του συντελεστή εσωτερικής συνάφειας Cronbach's alpha.²⁹
- Η συνολική τιμή του συντελεστή εσωτερικής συνάφειας Cronbach's alpha σχετίζεται θετικά με τις τιμές των επιμέρους συντελεστών Cronbach's alpha των παραγόντων της κλίμακας (από τις φορτίσεις τους), με τις μέσες συσχετίσεις μεταξύ των ερωτήσεων της κλίμακας, με το συστηματικό σφάλμα, αλλά και με το μέγεθος του δείγματος.^{28,30} Όσο μικρότερες είναι οι φορτίσεις των ερωτήσεων, τόσο λιγότερο αντιπροσωπευτική είναι η τιμή του συντελεστή εσωτερικής συνάφειας Cronbach's alpha. Όταν το συστηματικό σφάλμα είναι μικρό (0 και 0,1), τότε μεγαλύτερο δείγμα πληθυσμού δίνει αντιπροσωπευτικότερες τιμές του συντελεστή Cronbach's alpha. Συνεπώς, απαιτείται μεγάλο δείγμα πληθυσμού και μηδενικό συστηματικό σφάλμα. Σε μια πρόσφατη έρευνα, οι Shevlin et al τονίζουν ότι σε περίπτωση που έχουμε ορίσει ένα μέσο αποδεκτό όριο φόρτισης των παραγόντων (0,3), το συστηματικό σφάλμα εξακολουθεί να επηρεάζει την τιμή του συντελεστή Cronbach's alpha, χωρίς όμως να συμβαίνει το ίδιο και με το μέγεθος του δείγματος.³⁰

Κάποιοι ερευνητές προτείνουν ως αποδεκτούς τους ακόλουθους σταθμισμένους συντελεστές αξιοπιστίας:

- $a = 0,95$ για την εσωτερική συνοχή της κλίμακας

- r (Pearson coefficient) = 0,90 για επαναληπτική μέτρηση (test-retest)
- $a = 0,85$ για παραλλαγές της κλίμακας.

Επειδή οι αριθμοί αυτοί θεωρούνται υπερεκτιμημένοι, το όριο για το χαρακτηρισμό μιας κλίμακας ως αξιόπιστης^{31,32} είναι ένας συντελεστής $a > 0,7$. Ο Nunnally, ωστόσο, υποστηρίζει ότι ένας συντελεστής εσωτερικής συνάφειας Cronbach's alpha μεταξύ 0,5 και 0,6 είναι αρκετός στα αρχικά στάδια της μελέτης,³³ ενώ όταν πρόκειται να εξαχθούν σημαντικά συμπεράσματα, το ελάχιστο είναι ένας $a = 0,90$ με πιο επιθυμητή την τιμή $a = 0,95$.

Ιδιαίτερα σε ό,τι αφορά τη βιβλιογραφία για την ικανοποίηση των ασθενών, είναι κοινό φαινόμενο η αναφορά υψηλών συντελεστών εσωτερικής συνάφειας Cronbach's alpha. Ο Fitzpatrick εξηγεί ότι αυτό το φαινόμενο οφείλεται στο γεγονός πως οι ασθενείς τείνουν να εκφράζουν υψηλά επίπεδα ικανοποίησης από όλες τις διαστάσεις της παρεχόμενης φροντίδας, σε σημείο που να καθίσταται προβληματική η εμπιστοσύνη στις συσχετίσεις μεταξύ των ερωτήσεων, ως δείκτη αξιοπιστίας της αντίστοιχης κλίμακας.³⁴ Αν προσθέσουμε και το φαινόμενο του «φωτοστέφανου» (halo effect), ενισχύουμε την προηγούμενη διαπίστωση.

3.1. Τα είδη της αξιοπιστίας μιας κλίμακας

3.1.1. Η αξιοπιστία μεταξύ διαφορετικών βαθμολογητών και παρατηρητών (inter-rater, inter-observer reliability). Οποτεδήποτε χρησιμοποιούμε δύο ή περισσότερους βαθμολογητές ή παρατηρητές για την αξιολόγηση μιας δοκιμασίας, αναρωτιόμαστε αν και κατά πόσο τα αποτελέσματα είναι αξιόπιστα ή συνεπή. Αυτό είναι πολύ σημαντικό, ιδιαίτερα στις κλινικές και στις προβολικές δοκιμασίες. Όταν οι βαθμολογητές χρησιμοποιούν ονοματικές μετρήσεις, τότε υπολογίζεται ο συντελεστής συμφωνίας K (coefficient of agreement) ως δείκτης αξιοπιστίας.^{5,35} Είναι ο ιδανικός τρόπος εκτίμησης της αξιοπιστίας όταν χρησιμοποιούμε την παρατήρηση.

3.2.1. Αξιοπιστία επαναληπτικών μετρήσεων (test-retest reliability). Σύμφωνα με αυτή τη μέθοδο, επαναλαμβάνουμε την ίδια δοκιμασία στα ίδια άτομα, κάτω από τις ίδιες συνθήκες και ακολούθως αναζητούμε συσχέτιση ανάμεσα στα αποτελέσματα. Συνήθως, όσο μικρότερος είναι ο χρόνος που μεσολαβεί για τη δεύτερη μέτρηση, τόσο ισχυρή είναι η συσχέτιση.

3.1.3. Αξιοπιστία παράλληλων τύπων (parallel-forms reliability). Σε αυτή την περίπτωση, χορηγούμε έναν παράλληλο τύπο της κλίμακας μας στα ίδια άτομα, κάτω

από τις ίδιες συνθήκες και μετά αναζητούμε τη συσχέτισή τους. Ένας εναλλακτικός τρόπος είναι να δημιουργήσουμε μια μεγάλη ομάδα ερωτήσεων που αντιπροσωπεύουν την ίδια κατασκευή και μετά να χωρίσουμε τυχαία τις ερωτήσεις σε δύο κατηγορίες. Δίνουμε και τις δύο κατηγορίες ερωτήσεων στους ίδιους ανθρώπους. Ο συντελεστής συσχέτισης μεταξύ αυτών των δύο κατηγοριών ερωτήσεων θεωρείται δείκτης της αξιοπιστίας της κλίμακας. Ασφαλώς, η μέθοδος αυτή έχει πολλές δυσκολίες, καθώς είναι δύσκολο να δημιουργήσει κανείς ερωτήσεις που αντανακλούν την ίδια κατασκευή και να έχει δύο κατηγορίες παράλληλων ή ισοδύναμων ερωτήσεων. Για τους λόγους αυτούς, προτείνεται η χρήση του τύπου Α ή Β της κλίμακας στο πρωτογενές στάδιο της έρευνας και στη συνέχεια η ταυτόχρονη χορήγηση και των δύο στο τελικό στάδιο.

3.1.4. Αξιοπιστία εσωτερικής συνάφειας (internal consistency). Στην περίπτωση αυτή, χορηγούμε την κλίμακα σε ένα δείγμα ατόμων, μία φορά, προκειμένου να εκτιμήσουμε την αξιοπιστία της, δηλαδή πόσο καλά οι ερωτήσεις αντανακλούν την ίδια κατασκευή. Ο συντελεστής αξιοπιστίας Kuder-Richardson 20 και 21 (KR-20, KR-21) είναι ένας συντελεστής αξιοπιστίας που χρησιμοποιείται σε κλίμακες με απαντήσεις δυαδικής επιλογής του τύπου «ναι-όχι» και «συμφωνώ-διαφωνώ».³⁶

3.1.5. Αξιοπιστία των δύο ημίσεων ή των δύο ημίκλαστων (split-half reliability). Στην αξιοπιστία των δύο ημίσεων χορηγούμε τη δοκιμασία μία φορά σε ένα δείγμα του πληθυσμού και υπολογίζουμε τη συνολική βαθμολογία για κάθε ήμισυ τυχαία. Στην περίπτωση αυτή, ο δείκτης αξιοπιστίας των δύο ημίσεων είναι απλά ο συντελεστής συσχέτισης ανάμεσα στις δύο βαθμολογίες. Ο συντελεστής Cronbach's alpha είναι, από μαθηματικής σκοπιάς, ισοδύναμος με το μέσο όρο όλων των δυνατών εκτιμήσεων των ημίσεων.

3.1.6. Τι κλίμακα Likert πρέπει να χρησιμοποιήσω για να έχω έναν καλό συντελεστή αξιοπιστίας; Η χρήση κλιμάκων, όπως του τύπου Likert, αποτελεί μια κοινή πρακτική στις περισσότερες έρευνες. Οι Schutz και Rucker αναφέρουν ότι ο αριθμός (εύρος) των εναλλακτικών απαντήσεων δεν φαίνεται να επηρεάζει τα αποτελέσματα που προκύπτουν,³⁷ κάτι με το οποίο δεν συμφωνούν οι υπόλοιποι ερευνητές, με πιο ακραία την τοποθέτηση του Garner, ο οποίος προτείνει τη χρήση κλιμάκων με πάνω από 20 κατηγορίες απαντήσεων για την άντληση της μέγιστης δυνατής πληροφορίας.³⁸ Από την άλλη μεριά, οι Green και Rao προτείνουν τη χρήση 6βαθμων ή 7βαθμων κλιμάκων, πιστεύοντας ότι η αύξηση του εύρους των απαντήσεων πέραν της 7βαθμης κλίμακας δί-

νει λίγη παραπάνω πληροφορία.³⁹ Ο Symonds ήταν ο πρώτος που υποστήριξε τη θετική συσχέτιση της αξιοπιστίας με τη χρήση 7βαθμης κλίμακας απαντήσεων.⁴⁰ Αυτή η θέση ενδυναμώθηκε από τον Miller,⁴¹ ο οποίος υποστήριξε ότι *το ανθρώπινο μυαλό έχει τη δυνατότητα να διακρίνει ένα εύρος 7 απαντήσεων με μια απόκλιση ± 2* . Η χρήση της 7βαθμης κλίμακας προτάθηκε και από άλλους.^{29,42,43} Κάποιοι ερευνητές ανέφεραν υψηλότερους συντελεστές αξιοπιστίας για τις 5βαθμες κλίμακες.⁴⁴⁻⁴⁷

Αργότερα, ο Bendig διαπιστώνει μια σταθερότητα στους συντελεστές αξιοπιστίας επαναληπτικών μετρήσεων για τις κλίμακες Likert με 2, 3, 5, 7 και 9 απαντήσεις και μια σταθερότητα στους συντελεστές αξιοπιστίας μεταξύ των παρατηρητών/βαθμολογητών (inter-rater) για τις κλίμακες με 3, 5, 7 απαντήσεις και μείωση για τις 11βαθμες κλίμακες.^{48,49} Λίγοι μόνο ερευνητές συμφώνησαν με τον Bendig, ότι δηλαδή η αξιοπιστία είναι ανεξάρτητη από τον αριθμό των κατηγοριών των πιθανών απαντήσεων.⁵⁰⁻⁵⁵

Άλλοι ερευνητές χρησιμοποίησαν ως κριτήριο την εγκυρότητα, αναφέροντας ότι και η αξιοπιστία και η εγκυρότητα είναι ανεξάρτητες από το εύρος των εναλλακτικών απαντήσεων, ακόμη και όταν χρησιμοποιούμε κλίμακες 2 ή 3 σημείων.⁵³ Οι Loken et al, βασισμένοι στο κριτήριο της εγκυρότητας, βρήκαν ότι η 11βαθμη κλίμακα είναι ανώτερη της 3βαθμης και της 4βαθμης.⁵⁶ Οι Hancock και Klofars υποστήριξαν την 9βαθμη κλίμακα, έναντι της 5βαθμης.⁵⁷ Ο Chang βρήκε παρόμοιους συντελεστές εγκυρότητας με βάση το κριτήριο για τις 4βαθμες και 6βαθμες κλίμακες, αλλά καλύτερο συντελεστή συγκλίνουσας εγκυρότητας για την 6βαθμη κλίμακα.⁵⁸ Ο Jones μελέτησε τις προτιμήσεις των ανθρώπων για τη δυαδική ή 7βαθμη κλίμακα και κατέληξε στο συμπέρασμα ότι η δυαδική κλίμακα χαρακτηρίστηκε ως λιγότερο «ακριβής», λιγότερο «αξιόπιστη», λιγότερο «ενδιαφέρουσα» και περισσότερο «διφορούμενη» και «δύσκολη» από την 7βαθμη.⁵⁹ Σε μια πολύ πρόσφατη έρευνα των Preston και Colman,⁶⁰ στην οποία συμμετείχαν 149 φοιτητές ηλικίας 18–60 ετών, διαπιστώθηκε ότι η μικρότερη τιμή του συντελεστή Cronbach's alpha αντιστοιχεί στις δυαδικές και 3βαθμες κλίμακες ($\alpha=0,86-0,88$), ενώ η μέγιστη δυνατή, στην 7βαθμη, 8βαθμη, 9βαθμη και 10βαθμη κλίμακα ($\alpha=0,94$). Αντίστοιχα, η εγκυρότητα με βάση το κριτήριο μεγιστοποιείται με τη χρήση 7βαθμων, 8βαθμων ($\alpha=0,87$) και κυρίως 9βαθμων ($\alpha=0,89$) κλιμάκων. Η συγκλίνουσα εγκυρότητα είναι μεγαλύτερη στην 8βαθμη και την 9βαθμη κλίμακα. Οι συμμετέχοντες βρήκαν συνολικά πιο ευχάριστη και προτιμότερη τη 10βαθμη, την 7βαθμη και την 9βαθμη κλίμακα. Η διακρίνουσα δύναμη (discriminating power) υ-

πήρξε μεγαλύτερη στην 9βαθμη και τη 10βαθμη κλίμακα.

4. ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ (FACTOR ANALYSIS)

Στη σχετική βιβλιογραφία παρατίθεται μια σειρά από ορισμούς της παραγοντικής ανάλυσης, η οποία χρησιμοποιείται προκειμένου να ανιχνευθούν οι διαστάσεις μιας σειράς μεταβλητών. Με τη διαδικασία αυτή, ένας μεγάλος αριθμός μεταβλητών μειώνεται σε ένα μικρότερο αριθμό παραγόντων και ως εκ τούτου δεν περιγράφει μια εξαρτημένη μεταβλητή.^{61,62} Έτσι, η παραγοντική ανάλυση «σχεδιάστηκε για να εξετάσει τη συνδιακύμανση μιας ομάδας μεταβλητών και να ερμηνεύσει τις συσχετίσεις ανάμεσα σε αυτές τις μεταβλητές, ομαδοποιώντας τις σε παράγοντες».

Για τους Reymont και Joreskog, η παραγοντική ανάλυση «είναι ένας γενικός όρος, που χρησιμοποιούμε προκειμένου να περιγράψουμε μια σειρά από μεθόδους, οι οποίες σχεδιάστηκαν για να αναλύσουν τις σχέσεις μεταξύ κάποιων μεταβλητών ή αντικειμένων που άγουν στη διαμόρφωση υποθετικών μεταβλητών ή αντικειμένων, που ονομάζονται παράγοντες και υποτίθεται ότι περιέχουν την ίδια σημαντική πληροφορία που περιέχεται και σε μια μεγαλύτερη ομάδα παρατηρούμενων μεταβλητών ή αντικειμένων, τα οποία μειώνουν τη συνολική συνθετικότητα των δεδομένων».⁶³

Η διαδικασία της παραγοντικής ανάλυσης αναπτύχθηκε για πρώτη φορά στις αρχές του 20ού αιώνα από τον ψυχολόγο C. Spearman, ο οποίος υπέθεσε ότι μια ποικιλία δοκιμασιών νοητικής ικανότητας θα μπορούσαν να ερμηνευτούν από έναν «παράγοντα» της γενικής νοημοσύνης που είναι κοινός για όλες τις δοκιμασίες και τον χαρακτήρισε ως παράγοντα g.⁶⁴ Η ιδέα του υπήρξε ενδιαφέρουσα, αλλά αποδείχθηκε εσφαλμένη, αφού σίμερα πιστεύουμε σε 3 τουλάχιστον σημαντικούς παράγοντες της νοητικής ικανότητας: *προφορική, αριθμητική και λογική*. Λόγω της πολυπλοκότητάς της, άρχισε να χρησιμοποιείται ευρέως όταν ενσωματώθηκε σε στατιστικά πακέτα που παρέχονταν σε ηλεκτρονική μορφή.⁶⁵ Ο Kerlinger περιέγραψε την παραγοντική ανάλυση ως «ένα από τα πλέον δυνατά εργαλεία για τη μελέτη των πολύπλοκων πτυχών της ανθρώπινης συμπεριφοράς».⁶⁶

4.1. Είδη της παραγοντικής ανάλυσης

Μια τυπική παραγοντική ανάλυση καλείται να απατήσει σε 4 βασικά ερωτήματα: (α) πόσους διαφορετι-

κούς παράγοντες χρειαζόμαστε για να ερμηνεύσουμε τις σχέσεις μεταξύ των μεταβλητών; (β) ποια είναι η φύση αυτών των παραγόντων; (γ) πόσο καλά οι υποθετικοί παράγοντες ερμηνεύουν τα παρατηρούμενα δεδομένα; (δ) πόση τυχαία διακύμανση περιλαμβάνει κάθε παρατηρούμενη μεταβλητή;

4.4.1. Διερευνητική παραγοντική ανάλυση (exploratory factor analysis). Η διερευνητική παραγοντική ανάλυση χρησιμοποιείται για τη διερεύνηση των δεδομένων που προσδιορίζουν τον αριθμό ή τη φύση των παραγόντων που εξηγούν τη συνδιακύμανση μεταξύ των μεταβλητών, όταν ο ερευνητής δεν έχει εκ των προτέρων την κατάλληλη μαρτυρία για να σχηματίσει μια υπόθεση για τον αριθμό των παραγόντων που ερμηνεύουν τα δεδομένα. Ως εκ τούτου, αυτό το είδος της παραγοντικής ανάλυσης προσφέρει υποστήριξη στη διάχυση μιας θεωρίας, παρά στον έλεγχο αυτής της θεωρίας.⁶⁷ Ο Mulaik συμπεραίνει: «δεν υπάρχει καμιά αμφιβολία ότι η διερευνητική παραγοντική ανάλυση είναι χρήσιμη για τη διαμόρφωση υποθέσεων προς διερεύνηση, χωρίς αυτό να σημαίνει ότι είναι και η μοναδική μέθοδος. Με αυτή τη λογική, η απευθείας εμπειρία και επαφή κάποιου με ένα φαινόμενο είναι τις περισσότερες φορές ο απλούστερος τρόπος για τη διαμόρφωση αυτών των υποθέσεων».⁶⁸

Από τη στιγμή που μέσω της διαδικασίας της παραγοντικής ανάλυσης προσδιορίζεται αν και κατά πόσο οι παράγοντες σχετίζονται μεταξύ τους η διερευνητική παραγοντική ανάλυση βοηθά τους ερευνητές να τεκμηριώσουν την εγκυρότητα εννοιολογικής κατασκευής των ερωτήσεων της κλίμακας. Ας μην ξεχνάμε ότι η πλειονότητα των σημαντικών ερευνών είχαν διερευνητικό προφίλ.^{62,69} Αυτή η συσχέτιση διερευνητικής παραγοντικής ανάλυσης και εγκυρότητας εννοιολογικής κατασκευής αποτέλεσε σημείο αιχμής για τους στατιστικούς, με πρώτον τον Nunnally να υπογραμμίζει ότι «οι διερευνητικές μέθοδοι δεν συνιστούν ένα βασιλικό δρόμο προς την αλήθεια».³² Από εκεί και πέρα ακολούθησαν και άλλες κριτικές, όπως του Mulaik: «δεν υπάρχουν προφανείς τρόποι για να αντλήσουμε γνώση από την εμπειρία, παρά μόνο αν κάνουμε εκ των προτέρων ορισμένες παραδοχές».⁷⁰ Από την άλλη μεριά, η διερευνητική παραγοντική ανάλυση ακολουθεί ένα γραμμικό μοντέλο κατάλληλο για συγκεκριμένη κατηγορία δεδομένων, με αποτέλεσμα να ερμηνεύει με εσφαλμένο τρόπο μη γραμμικές σχέσεις δεδομένων. Οι τεχνικές της διερευνητικής παραγοντικής ανάλυσης δεν μας δίνουν μια ένδειξη σφάλματος των απόψεων κάποιου, από τη στιγμή που η τεχνική σχεδιάστηκε για να ταιριάζει τα δεδομένα. Έτσι, μπορεί να προτείνει κάποιες υποθέσεις, ωστόσο δεν επικυρώνει

τη γνώση. Αυτό μοιραία οδηγεί σε μια δυσκολία ερμηνείας των ίδιων των παραγόντων που προκύπτουν.

4.1.2. Επιβεβαιωτική παραγοντική ανάλυση (confirmatory factor analysis). Η επιβεβαιωτική παραγοντική ανάλυση συνιστά ένα μοναδικό τρόπο για τον έλεγχο μιας θεωρίας, καθώς ο ερευνητής ξεκινά με μια υπόθεση πριν από την ανάλυση. Η υπόθεση αυτή βασίζεται σε μια ισχυρή θεωρία ή μια παρατήρηση, ενώ ορίζει ποιες μεταβλητές θα σχετίζονται με ποιους παράγοντες, όπως επίσης και ποιοι παράγοντες θα σχετίζονται μεταξύ τους.⁶⁷ Εξ ορισμού, λοιπόν, αυτού του είδους η παραγοντική ανάλυση είναι πιο αξιόπιστη στην αξιολόγηση της εγκυρότητας εννοιολογικής κατασκευής της κλίμακας. Έχοντας προσδιορίσει ο ερευνητής εκ των προτέρων τους παράγοντες, είναι σε θέση να τους συγκρίνει με αυτούς που βρέθηκαν από μια ομάδα δεδομένων, προκειμένου να προσδιορίσει την καταλληλότητα (goodness of fit) του μοντέλου. Στην επιβεβαιωτική παραγοντική ανάλυση, ο ερευνητής προτείνει κάποια μοντέλα, που βασίζονται στη θεωρία ή τα υπάρχοντα δεδομένα, τα οποία υποτίθεται ότι ταιριάζουν στα δεδομένα. Στα μοντέλα αυτά, προσδιορίζονται οι συσχετίσεις των παραγόντων και των μεταβλητών με έναν ή περισσότερους παράγοντες, σύμφωνα με τις θεωρητικές προσδοκίες του ερευνητή, αφήνοντας και κάποιες παραμέτρους να τις προσδιορίσει η ίδια η ανάλυση.⁷¹ Η επιβεβαιωτική παραγοντική ανάλυση μπορεί να γίνει με τη βοήθεια προγραμμάτων, όπως το LISREL, μέσω του οποίου ελέγχονται τα θεωρητικά μοντέλα.⁷² Αυτές οι στατιστικές μέθοδοι αναφέρονται ως στατιστικές ελέγχου της καταλληλότητας (fit statistics) και συμβάλλουν στη διασφάλιση εκείνου του μοντέλου που εξηγεί καλύτερα τις σχέσεις ανάμεσα στις παρατηρούμενες και τις ορισθείσες μεταβλητές.⁷³

Για την επιβεβαιωτική παραγοντική ανάλυση δεν υπάρχει όριο στον αριθμό των μεταβλητών που εισάγονται. Για τη διερευνητική, ο Thurstone⁷⁴ προτείνει τουλάχιστον 3 μεταβλητές ανά παράγοντα.⁶⁹ Η άποψη όσο περισσότερες μεταβλητές τόσο το καλύτερο για την παραγοντική ανάλυση, δεν αληθεύει. Έτσι, ερωτήσεις –σε μια κλίμακα– του τύπου: «μου αρέσει το γραφείο μου», «το γραφείο μου είναι ωραίο», «μου αρέσει να δουλεύω στο γραφείο μου», «νιώθω ωραία όταν δουλεύω στο γραφείο μου» μπορούν να διαμορφώσουν τον παράγοντα *γραφείο*, τη στιγμή που ο ερευνητής αναζητά τον ευρύτερο παράγοντα *ικανοποίηση από την εργασία*. Έτσι, θα πρέπει να ξεκινά κανείς με μια μικρή ομάδα ερωτήσεων με την υψηλότερη εγκυρότητα περιεχομένου.

4.2. Μορφές της παραγοντικής ανάλυσης

4.2.1. Παραγοντική ανάλυση κυρίων συνιστωσών (principal factor analysis). Είναι η πιο συχνή μορφή παραγοντικής ανάλυσης που χρησιμοποιεί γραμμικό συνδυασμό των μεταβλητών, τέτοιον ώστε να εξάγεται η μεταβλητή με τη μεγαλύτερη διακύμανση. Μετά την εξαγωγή της συγκεκριμένης μεταβλητής ακολουθεί ένας δεύτερος γραμμικός συνδυασμός, ο οποίος εξηγεί τη μέγιστη ποσοστιαία διακύμανση που απομένει κ.ο.κ. Αυτή η μέθοδος καλείται και *μέθοδος κύριου άξονα* (principal axis method) και οδηγεί σε ορθογώνιους (μη σχετιζόμενους) παράγοντες.

4.2.2. Κανονική παραγοντική ανάλυση (canonical factor analysis). Καλείται και κανονική παραγοντική ανάλυση του Rao (Rao's canonical factor analysis). Είναι μια διαφορετική μέθοδος υπολογισμού του ίδιου μοντέλου με την παραγοντική ανάλυση κυρίων συνιστωσών. Με τη μέθοδο αυτή, αναζητούνται οι παράγοντες που εμφανίζουν την υψηλότερη κανονική συσχέτιση με τις μεταβλητές.

4.2.3. Ανάλυση κύριων παραγόντων (principal factor analysis). Καλείται και παραγοντική ανάλυση κύριων αξόνων (principal axis factoring), αλλά και κοινή παραγοντική ανάλυση. Πρόκειται για μια μορφή παραγοντικής ανάλυσης που αναζητά το μικρότερο αριθμό παραγόντων που ερμηνεύουν την κοινή διακύμανση των μεταβλητών.

4.2.4. R-mode παραγοντική ανάλυση. Με αυτή τη μέθοδο της παραγοντικής ανάλυσης, οι γραμμές είναι οι μεταβλητές (ερωτήσεις της κλίμακας), οι στήλες είναι οι περιπτώσεις και τα κελιά είναι η βαθμολογία των περιπτώσεων στις αντίστοιχες μεταβλητές.

4.2.5. Q-mode παραγοντική ανάλυση. Στην περίπτωση αυτή, οι γραμμές είναι οι περιπτώσεις και οι στήλες είναι οι μεταβλητές. Ένα πρόβλημα με αυτού του είδους την παραγοντική ανάλυση είναι οι αρνητικές φορτίσεις στους παράγοντες που δεν έχουν πάντα μια σαφή ερμηνεία.

4.3. Κριτήρια για τον προσδιορισμό των παραγόντων

Η αναζήτηση των παραγόντων γίνεται με βάση κάποια κριτήρια,⁷⁵ τα οποία –σύμφωνα με τη συχνότητα χρήσης στις κοινωνικές επιστήμες– είναι:

- **Το κριτήριο Kaiser:** Ο κανόνας του Kaiser⁷⁶ προτείνει την απόρριψη των παραγόντων με ιδιοτιμή (eigenvalue) μικρότερη του 1.
- **Διάγραμμα ιδιοτιμών (scree plot ή scree test):** Διατυπώθηκε από τον Cattell⁷⁷ και απεικονίζει την κατανο-

μή των ιδιοτιμών ως προς τους παράγοντες. Προτείνει την αποδοχή τόσων παραγόντων όσων αντιστοιχούν σε ιδιοτιμή μεγαλύτερη του 1.

- **Το ποσοστό της συνολικής διακύμανσης,** που ερμηνεύουν οι παράγοντες και το οποίο θεωρείται σημαντικό όταν είναι μεγαλύτερο από το 80%. Στην ουσία, όμως, στην πλειοψηφία των ερευνών αυτό είναι πάρα πολύ δύσκολο να επιτευχθεί.
- **Το κριτήριο Joliffe:** Πρόκειται για έναν κανόνα που χρησιμοποιείται ελάχιστα και προτείνει ως αποδεκτές παραγοντικές φορτίσεις όσες είναι πάνω από 0,70, που κατά άλλους θεωρούνται πολύ υψηλές.⁷⁸
- **Το είδος της στρέψης:** Η στρέψη των αξόνων καθιστά πιο εύκολη την ερμηνεία των παραγόντων. Το άθροισμα των ιδιοτιμών δεν επηρεάζεται από το είδος της στρέψης, δεν συμβαίνει όμως το ίδιο με τις ιδιοτιμές συγκεκριμένων παραγόντων:
 - **Καμιά στρέψη.** Η ανάλυση κυρίων συνιστωσών, χωρίς στρέψη των αξόνων, μεγιστοποιεί το άθροισμα των τετραγώνων των παραγοντικών φορτίσεων, διαμορφώνοντας έναν αριθμό παραγόντων που εξηγούν τη μεγαλύτερη δυνατή διακύμανση, δυσκολεύοντας όμως την ερμηνεία των παραγόντων, καθώς οι μεταβλητές τείνουν να φορτίζουν πολλούς παράγοντες ταυτοχρόνως.
 - **Varimax στρέψη (variance maximizing rotation).** Είναι η συχνότερα χρησιμοποιούμενη μέθοδος στρέψης των αξόνων. Ονομάζεται έτσι, διότι ο στόχος μας από τη στρέψη είναι η μεγιστοποίηση της διακύμανσης της νέας μεταβλητής, που είναι ο παράγοντας, μειώνοντας τη διακύμανση γύρω από το νέο παράγοντα. Πρόκειται για μια ορθογώνια στρέψη, που στοχεύει στη μεγιστοποίηση της διακύμανσης των τετραγώνων των φορτίσεων ενός παράγοντα (στήλη) για όλες τις μεταβλητές (γραμμή) σε έναν πίνακα φόρτισης παραγόντων, διαφοροποιώντας τις γνήσιες μεταβλητές από τους εξαχθέντες παράγοντες. Αυτό πρακτικά σημαίνει ότι μειώνεται ο αριθμός των μεταβλητών που παρουσιάζουν υψηλές φορτίσεις σε κάθε έναν παράγοντα. Αποτελεί τη μέθοδο εκλογής για όσους ζητούν ερμηνεία των παραγόντων τους.
 - **Quartimax στρέψη.** Είναι μια ορθογώνια στρέψη, που μειώνει τον αριθμό των απαιτούμενων παραγόντων για την ερμηνεία κάθε μεταβλητής. Είναι η μέθοδος που χαρακτηρίζεται ως εκκαθάριση των μεταβλητών⁶⁷ (clean up the variables), αφού αναδεικνύει ένα γενικό παράγοντα, επειδή η κάθε με-

ταβλητή τείνει να σχετίζεται πρωταρχικά με έναν παράγοντα.

- *Equimax στρέψη*. Είναι μια ενδιάμεση μέθοδος μεταξύ της *varimax* και *quartimax* στρέψης.
- *Direct oblimin στρέψη*. Είναι η μέθοδος εκλογής για όσους δεν επιθυμούν μια μη ορθογώνια λύση. Ουσιαστικά, μέσω αυτής της μεθόδου καταλήγουμε σε υψηλότερες ιδιοτιμές, αλλά σε χαμηλή ερμηνεία των παραγόντων.
- *Promax στρέψη*. Είναι μια εναλλακτική μη ορθογώνια στρέψη, η οποία χρησιμοποιείται όταν υπάρχουν πολλά δεδομένα.

Φορτίσεις. Σε γενικές γραμμές, παραγοντικές φορτίσεις της τάξης του 0,71 θεωρούνται πολύ υψηλές, 0,63 υψηλές, 0,55 ικανοποιητικές, 0,45 μέτριες, 0,32 χαμηλές και κάτω από 0,30 μη αξιολογήσιμες.⁷⁸ Οι περισσότεροι ερευνητές διατείνονται ότι δεν υπάρχει κάποια ομοφωνία στις τιμές, ωστόσο συμφωνούν στο κατώφλι 0,30 ή 0,35 ως τη μικρότερη αποδεκτή τιμή φόρτισης. Οι Norman και Streiner δίνουν έναν εναλλακτικό τύπο για τον υπολογισμό της ελάχιστης αποδεκτής φόρτισης για δείγμα μεγαλύτερο ή ίσο των 100 ατόμων.⁷⁹

Δεν υπάρχει επιστημονική απάντηση στην ερώτηση πόσες περιπτώσεις απαιτούνται για την εκτέλεση παραγοντικής ανάλυσης, καθώς οι απόψεις δίστανται. Από την ανασκόπηση της σχετικής βιβλιογραφίας προκύπτουν τα εξής:

- *Ο κανόνας των 10*. Θα πρέπει να υπάρχουν τουλάχιστον 10 περιπτώσεις για κάθε μία ερώτηση της κλίμακας που χρησιμοποιείται.³³ Αυτό ισχυρίζεται και η Hultka για την ικανοποίηση των ασθενών.⁸⁰
- *Αναλογία ατόμων/μεταβλητές*. Η αναλογία ατόμων προς μεταβλητές⁸¹ δεν πρέπει να είναι μικρότερη από 5.
- *Ο κανόνας των 100*. Ο αριθμός των ατόμων του δείγματος πρέπει να είναι 5 φορές μεγαλύτερος του αριθμού των μεταβλητών ή να είναι 100. Ίσως χρειαστούν περισσότερα άτομα, αν οι κοινότητες διακύμανσης των παραγόντων είναι μικρές και/ή λίγες μόνο μεταβλητές φορτίζουν σε κάθε παράγοντα.⁸²
- *Ο κανόνας των 150*. Οι Hutcheson και Sofroniou προτείνουν τουλάχιστον 150–300 περιπτώσεις.⁸³
- *Ο κανόνας των 200*. Πρέπει να υπάρχουν τουλάχιστον 200 περιπτώσεις, ανεξάρτητα από την αναλογία ατόμων προς μεταβλητές.⁶²

Ο Ware έδωσε μια ερμηνεία για το ποσοστό της διακύμανσης που εξηγεί η παραγοντική ανάλυση στις κλίμακες για την εκτίμηση της ικανοποίησης των ασθενών: «μεταξύ των ασθενών με λιγότερα χρόνια εκπαίδευσης υπάρχει η τάση να συμφωνούν με όλες τις ερωτήσεις της κλίμακας, κάτι που επηρεάζει –θετικά ή αρνητικά– ακόμη και σε ποσοστό 25% τη διακύμανση στους πίνακες συσχετίσεων μεταξύ των ερωτήσεων».⁸⁴ Έτσι, μειώνεται η δυνατότητα ερμηνείας των ίδιων των παραγόντων. Ένας επιπλέον «συγχυτικός» παράγοντας ερμηνείας είναι η αδυναμία των ασθενών να διαχωρίσουν τη φροντίδα (care) από τη θεραπεία (cure).

4.4. Επάρκεια της δειγματοληψίας

Η επάρκεια της δειγματοληψίας (sampling adequacy) εκτιμάται με το δείκτη Kaiser-Meyer-Olkin (KMO).⁷⁶ Υπάρχει ένας KMO για κάθε μεταβλητή, που κυμαίνεται από 0 έως 1,0. Η αποδεκτή τιμή του δείκτη KMO για την πραγματοποίηση μιας παραγοντικής ανάλυσης είναι μεγαλύτερη από 0,6.

4.5. Στρέψη των αξόνων

Για την περιστροφή των παραγόντων απαιτείται μια σαφής μεθοδολογία.⁶² Κατά καιρούς έχουν προταθεί διάφορα κριτήρια, με προεξάρχοντα αυτά του Thurstone:⁷⁴ (α) κάθε μεταβλητή θα πρέπει να έχει έναν τουλάχιστον μηδενικό συντελεστή φόρτισης του παράγοντα, (β) κάθε παράγοντας πρέπει να διαθέτει μια σειρά μεταβλητών, των οποίων οι συντελεστές φόρτισης του παράγοντα είναι μηδενικοί, (γ) για κάθε ένα ζεύγος παραγόντων, μια μεγάλη αναλογία των μεταβλητών πρέπει να έχει μηδενικούς συντελεστές φόρτισης και στους δύο παράγοντες, κάθε φορά που προκύπτουν 4 παράγοντες και (δ) για κάθε ζεύγος παραγόντων, θα πρέπει να υπάρχει μόνο ένας μικρός αριθμός μεταβλητών με μη μηδενικούς συντελεστές.

Ο κυριότερος λόγος για τον οποίο ο Thurstone πρότεινε αυτά τα κριτήρια για τη στρέψη των αξόνων των παραγόντων είναι το γεγονός ότι η δομή του παράγοντα πρέπει να παραμένει η ίδια, ανεξάρτητα από τον αριθμό των μεταβλητών που τον περιγράφουν. Όταν βέβαια υπάρχει μόνο ένας παράγοντας, η στρέψη είναι περιττή. Οι τρόποι στρέψης των αξόνων χωρίζονται σε δύο κατηγορίες: την πλάγια (oblique) ή μη ορθογώνια και την ορθογώνια (orthogonal) στρέψη. Η πλάγια στρέψη χρησιμοποιείται πολύ σπάνια στις κοινωνικές έρευνες, γιατί, παρόλο που καθιστά πιο ευδιάκριτες τις σχέσεις μεταβλητών και παραγόντων, κάνει πολύ δύσκολη τη διάκρι-

ση μεταξύ των παραγόντων. Χρησιμοποιείται, ωστόσο, στην επιβεβαιωτική παραγοντική ανάλυση, στην περίπτωση που η θεωρία υποδεικνύει δύο παράγοντες που σχετίζονται μεταξύ τους. Στη διερευνητική παραγοντική ανάλυση η ορθογώνια στρέψη συνιστάται, επειδή έχει το πλεονέκτημα της απλότητας⁸⁵ και επειδή τα αποτελέσματα είναι περισσότερο ερμηνεύσιμα και εφαρμόσιμα στο μέλλον, χωρίς αυτό να σημαίνει ότι είναι κατ' ανάγκη αντιπροσωπευτική της πραγματικότητας, κάτι που συνιστά ένα μειονέκτημα της μεθόδου. Ωστόσο, τα οφέλη της ορθογώνιας στρέψης υπερσκελίζουν τις αδυναμίες της. Εάν η πλάγια στρέψη των αξόνων δίνει αποτελέσματα που *ταιριάζουν καλύτερα στα δεδομένα* του δείγματος, τότε η ορθογώνια στρέψη δίνει αποτελέσματα που *ταιριάζουν καλύτερα στα δεδομένα των ερευνών* που προηγήθηκαν ή πρόκειται να πραγματοποιηθούν στο μέλλον. Αυτό συμβαίνει διότι υπάρχει μικρότερο σφάλμα δειγματοληψίας στην ορθογώνια στρέψη. Οι Pedhazur και Schmelkin, για λόγους εγκυρότητας της εννοιολογικής κατασκευής, προτείνουν την ταυτόχρονη χρήση της ορθογώνιας και της πλάγιας στρέψης των αξόνων. Αν

στην πορεία προκύψει ότι οι συσχετίσεις μεταξύ των παραγόντων είναι αμελητέες, τότε υποστηρίζεται η ερμηνεία των αποτελεσμάτων της ορθογώνιας στρέψης. Ακόμη όμως και στην αντίθετη περίπτωση, η ορθογώνια στρέψη είναι απλούστερη.

Στην παραγοντική ανάλυση, το ποσοστό της διακύμανσης που οφείλεται σε κοινούς παράγοντες καλείται κοινότητα διακύμανσης των παραγόντων⁸⁶ (ΚΔΠ) (*communality*). Πρόκειται για το άθροισμα των τετραγώνων των φορτίσεων των παραγόντων για μια δεδομένη μεταβλητή (γραμμή) που αντιστοιχεί στη διακύμανση σε αυτή τη μεταβλητή, η οποία ερμηνεύεται από όλους τους παράγοντες. Μια επιπλέον ενέργειά μας, λοιπόν, όταν εκτελούμε παραγοντική ανάλυση, θα πρέπει να είναι η εκτίμηση των ΚΔΠ για κάθε ερώτηση της κλίμακας, δηλαδή του ποσοστού της διακύμανσης που είναι κοινή για κάθε ερώτηση με τις υπόλοιπες ερωτήσεις. Οι ΚΔΠ ερμηνεύονται πάντα σε συνάρτηση με την ερμηνεία των παραγόντων. Έτσι, μια ΚΠΔ 0,75 φαίνεται μεγάλη, εφόσον ο παράγοντας στον οποίο φορτίζει η

ABSTRACT

Scale validation methodology

V. RAFTOPOULOS,¹ H. THEODOSOPOULOU²

¹TEI, Larissa, ²Department of Nursing, University of Athens, Athens, Greece

Archives of Hellenic Medicine 2002, 19(5):577-589

The systematic use of psychometric scales, in psychology and psychiatry, but also in other research initiatives, necessitates the development and use of research methodology that assures validity and reliability of these scales. The aim of this paper is to describe such methodology, using examples from patient satisfaction literature.

Key words: Factor analysis, Patient satisfaction, Reliability, Scale validation, Validity

μεταβλητή ερμηνεύεται. Μια ΚΠΔ 0,25 ίσως μοιάζει μικρή, ωστόσο έχει αξία αν συναρτάται με έναν ερμηνεύσιμο παράγοντα. Συνεπώς, ο παράγοντας είναι εκείνος που προσδιορίζει τα χαρακτηριστικά της τιμής της ΚΠΔ.

Βιβλιογραφία

1. CRONBACH LJ. *Essentials of psychological testing*. 4th ed. New York, Harper & Row, 1984
2. HAMMERSLEY M. Some notes on the terms «validity» and «reliability». *British Educational Research Journal* 1987, 13:73-81
3. SIMCO N, WARIN J. Validity in image based research. An elaborated illustration of the issues. *British Educational Research Journal* 1997, 23:661-673
4. AMERICAN PSYCHOLOGICAL ASSOCIATION. *Ethical standards of*

psychologists. Washington, DC, American Psychological Association, Inc, 1953

5. ANASTASI A. The concept of validity in the interpretation of test scores. *Education Psychology Measurement* 1950, 10:67-78
6. CHURCHILL GA. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 1979, 16:64-73
7. PETER JP. Reliability: a review of psychometric basics and recent marketing practices. *Journal of Marketing Research* 1981, 16:6-17
8. CAMPBELL DT. Recommendations for APA test standards regarding construct, trait and discriminant validity. *Am Psychol* 1960, August:546-553

9. SMITH AM. Some problems when adopting Churchill's paradigm for the development of service quality measurement scales. *Journal of Business Research* 1999, 46:109–120
10. METERKO M, NELSON EC, RUBIN HR. Patients judgments of hospital quality. *Med Care* 1990, 28(Suppl 9):S1–S56
11. CARMINES EG, ZELLER RA. *Reliability and validity assessment*. Sage Publications Series Number 07-017. Newbury Park, CA, Sage Publ, Inc, 1979
12. CAMPBELL DT, FISKE DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959, 56:81–105
13. MARSH HW, SMITH ID. Multitrait-multimethod analyses of two self-concept instruments. *J Educ Psychol* 1982, 74:430–440
14. WYLIE RC. *Measures of self-concept*. Lincoln, NE, University of Nebraska Press, 1989
15. HALL JA, DORNAN MC. Meta-analysis of satisfaction with medical care: description of research domain and analyses of overall. *Soc Sci Med* 1988, 27:637–644
16. SITZIA J, WOOD N. Response rate in patient satisfaction research: an analysis of 210 published studies. *Int J Qual Health Care* 1998, 10:311–317
17. VAN MAANEN HMT. Improvement of quality of nursing care: a goal to challenge in eighties. *J Adv Nurs* 1981, 6:3–9
18. KIRK J, MILLER M. *Reliability and validity in qualitative research*. Qualitative Research Methods Series, No 1, London, Sage 9, 1986
19. CHENITZ WC, SWANSON JM. Qualitative research using grounded theory. In: *From practice to grounded theory: qualitative research in nursing*. Addison Wesley, Menlo Park, California, 1986:3–15
20. FISKE ST, TAYLOR SE. *Social Cognition*. New York, Random House, 1984
21. STRASSER S. *Measuring patient satisfaction for improved patient services*. Management Series, American College of Healthcare Executives, 1990
22. ROWLEY GL. The reliability of observational measures. *American Educational Research Journal* 1976, 13:51–59
23. CROCKER LM, ALGINA J. *Introduction to classical and modern test theory*. New York, Holt, Rinehart & Winston, 1986:144
24. GRONLUND NE, LINN RL. *Measurement and evaluation in teaching*. 6th ed. New York, MacMillan, 1990:78
25. VACHA-HAASE T. Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement* 1998, 58:6–20
26. CRONBACH A. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951, 16:297–334
27. BOLLEN KA. *Structural equations with latent variables*. New York, Wiley, 1989
28. CORTINA JM. What is coefficient alpha: an examination of theory and applications. *J Appl Psychol* 1993, 78:98–104
29. NUNNALLY JC, BERNSTEIN IH. *Psychometric theory*. 3rd ed. New York, McGraw-Hill, 1994
30. SHELVIN M, MILES JNV, DAVIES MNO, WALKER S. Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences* 2000, 28:229–237
31. NUNNALLY J. *Psychometric theory*. New York, McGraw-Hill, 1978
32. PAYNE B. 2b or not 2b? *Br Acad Management Newsletter* 1993, June:7
33. NUNNALLY JC. Cited in: Churchill GA Jr. Paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 1979, XVI (February)
34. FITZPATRICK R. Scope and measurement of patient satisfaction. In: Fitzpatrick R, Hopkins A (eds) *Measurement of patients' satisfaction with their care*. London, Royal College of Physicians, 1993
35. COHEN BJ, SWERDLIK ME, PHILLIPS SM. *Psychological testing and assessment: an introduction to tests and measurement*. 3rd ed. Mountain View, CA, Mayfield Publ Co, 1996
36. KUDRE GF, RICHARDSON MW. The theory of the estimation of test reliability. *Psychometrika* 1937, 2:151–160
37. SCHUTZ HG, RUCKER MH. A comparison of variable configurations across scale lengths: an empirical study. *Educational and Psychological Measurement* 1975, 35:319–324
38. GARNER WR. Rating scales, discriminability and information transmission. *Psychol Rev* 1960, 67:343–352
39. GREEN PE, RAO VR. Rating scales and information recovery. How many scales and response categories to use? *Journal of Marketing* 1970, 34:33–39
40. SYMONDS PM. On the loss of reliability in ratings due to coarseness of the scale. *J Exp Psychol* 1924, 7:456–461
41. MILLER GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956, 63:81–97
42. FINN RH. Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement* 1972, 34:885–892
43. RAMSAY JO. The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika* 1973, 38:513–533
44. JENKINS GD, TABER TD. A Monte-Carlo study of factors affecting three indices of composite scale reliability. *J Appl Psychol* 1977, 62:392–398
45. LISSITZ RW, GREEN SB. Effect of the number of scale points on reliability: a Monte-Carlo approach. *J Appl Psychol* 1975, 60:10–13
46. MCKELVIE SJ. Graphic rating scales: how many categories? *Br J Psychol* 1978, 69:185–202
47. REMMERS HH, EWART E. Reliability of multiple-choice measuring instruments as a function of the Spearman-Brown prophecy formula. *J Educ Psychol* 1941, 32:61–66
48. BENDIG AW. The reliability of self-ratings as a function of the amount of verbal anchoring and the number of categories on the scale. *J Appl Psychol* 1953, 37:38–41
49. BENDIG AW. Reliability and the number of rating scale categories. *J Appl Psychol* 1954, 38:38–40
50. BOOTE AS. Reliability testing of psychographic scales: five-point or seven-point? Anchored or labelled? *Journal of Advertising Research* 1981, 21:53–60
51. BROWN G, WILDING RE, COULTER RL. Customer evaluation of retail salespeople using the SOCO scale: a replication extension and application. *Journal of the Academy of Marketing*

- Science* 1991, 9:347–351
52. KOMORITA SS. Attitude content, intensity and the neutral point on a Likert scale. *J Soc Psychol* 1963, 61:327–334
 53. MATELL MS, JACOBY J. Is there an optimal number of alternatives for Likert scale items? Study I: reliability and validity. *Educational and Psychological Measurement* 1971, 31:657–674
 54. PEABODY D. Two components in bipolar scales: direction and extremeness. *Psychol Rev* 1962, 69:65–73
 55. REMINGTON M, TYRER PJ, NEWSON-SMITH J, CICHETTI DV. Comparative reliability of categorical and analogue rating scales in the assessment of psychiatric symptomatology. *Psychol Med* 1979, 9:765–770
 56. LOKEN B, PIRIE P, VIRNIG KA, HINKLE RL, SALMON CT. The use of 0–10 scales in telephone surveys. *Journal of Market Research Society* 1987, 29:353–362
 57. HANCOCK GR, KLOCKARS AJ. The effect of scale manipulations on validity: targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics* 1991, 22:147–154
 58. CHANG L. A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement* 1994, 18:205–215
 59. JONES RR. Differences in response consistency and subjects' preferences for three personality inventory response formats. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, pp. 247–248
 60. PRESTON CC, COLMAN AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power and respondent preferences. *Acta Psychol* 2000, 104:1–15
 61. CARR SC. A primer on the use of Q sort technique factor analysis. *Measurement and Evaluation in Counseling and Development* 1992, 25:133–138
 62. GORSUCH RL. *Factor analysis*. Hillsdale, NJ, Erlbaum, 1983
 63. REYMONT R, JORESKOG KG. *Applied factor analysis in the natural sciences*. New York, Cambridge University Press, 1993
 64. SPEARMAN C. "General intelligence" objectively determined and measured. *Am J Psychol* 1904, 15:201–293
 65. THOMPSON B, DENNINGS B. The unnumbered graphic scale as a data-collection method: investigation comparing three measurement strategies in the context of Q-technique factor analysis. Paper presented at the annual meeting of the Mid-South Educational Research Association. New Orleans. (ERIC Document Reproduction Service No ED 364589), 1993
 66. KERLINGER FN. *Foundations of behavioral research*. 3rd ed. New York, Holt, Rhinehart and Winston, 1986
 67. STEVENS J. *Applied multivariate statistics for the social sciences*. 3rd ed. Mahwah, NJ, Lawrence Erlbaum Associates, 1996
 68. MULAİK SA. *The foundations of factor analysis*. New York, McGraw Hill, 1972
 69. KIM JO, MUELLER CW. *Introduction to factor analysis*. Beverly Hills, Sage Publ, 1978
 70. MULAİK SA, JAMES LR, VAN ALSTINE J, BENNET N, LIND, STILWELL CD. Evaluation of goodness-of-fit indices for structural equation models. *Psychol Bull* 1989, 105:430–455
 71. GILLAPSY JA. *A primer on confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, 1996
 72. JORESKOG KG, SORBOM D. *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods*. 4th ed, 1986
 73. BENTLER PM. Multivariate analysis with latent variables: causal modeling. *Annu Rev Psychol* 1980, 31:11–21
 74. THURSTONE LL. *Multiple-factor analysis*. Chicago, Chicago University Press, 1947
 75. DUNTEMAN GH. *Principal component analysis*. Thousand Oaks, CA, Sage Publications, Quantitative applications in the Social Sciences Series No 69, 1989
 76. KAISER HF, RICE JL. *Educational and Psychological Management* 1974, 34:111–117
 77. CATTELL RB. The scree test for the number of factors. *Multivariate Behavioral Research* 1966, 1:245–276
 78. COMREY AL. *A first course in factor analysis*. New York, Academic Press, 1973
 79. NORMAN GR, STREINER L. *Biostatistics: the bare essentials*. St. Louis, MO, Mosby, 1994
 80. HULKA BS, ZYZANSKI SJ, CASSEL JC, THOMPSON SJ. Satisfaction with medical care in low income population. *J Cron Dis* 1971, 24:661–673
 81. BRYANT, YARNOLD. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm, Yarnold (eds) *Reading and understanding multivariate analysis*. American Psychological Association Books, 1995
 82. HATCER L. *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC, SAS Institute, Focus on the CALIS procedure, 1994
 83. HUTCHSON G, SOFRONIOU N. *The multivariate social scientist: introductory statistics using generalized linear models*. Thousand Oaks, CA, Sage Publ, 1999
 84. WARE JJE, SNYDER MK. Dimensions of patient attitudes regarding doctors and medical care services. *Med Care* 1975, 13:669–682
 85. PEDHAZUR EJ, SCHMELIKN LP. *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ, Erlbaum, 1991
 86. ΑΛΕΞΟΠΟΥΛΟΣ ΔΣ. *Ψυχομετρία. Σχεδιασμός τεστ και ανάλυση ερωτήσεων*. Α τόμος. Ελληνικά Γράμματα, 1998

Corresponding author:

V. Raftopoulos, 3 Makelari street, GR 142 32 Athens, Greece
E-mail: vraftop@nurs.uoa.gr