

Η ΕΓΚΥΡΟΤΗΤΑ ΚΑΙ Ο ΕΛΕΓΧΟΣ ΤΗΣ ΣΤΗΝ ΠΙΣΤΟΠΟΙΗΣΗ ΤΗΣ ΓΛΩΣΣΟΜΑΘΕΙΑΣ

Αντώνης Τσοπάνογλου
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Εισαγωγή

Κάποιοι αναγνώστες¹ αυτού του κειμένου ίσως αναρωτηθούν για ποιο λόγο φιλοξενείται εδώ, δηλαδή σε έναν ιστότοπο που ασχολείται με ζητήματα διδασκαλίας της αγγλικής ως ξένης γλώσσας και αξιολόγησης των ικανοτήτων χρήσης της, με τελικό στόχο την πιστοποίηση. Ένα άρθρο αφιερωμένο στην *εγκυρότητα*² (validity) και στις διαδικασίες ή μεθόδους *ελέγχου*³ της *εγκυρότητας* (validation), δύο έννοιες που λογικά αποτελούν θεωρητικά εργαλεία της μεθοδολογίας της έρευνας έχει τη θέση του εδώ επειδή υιοθετείται η θέση πως η πιστοποίηση της γλωσσομάθειας είναι μια δραστηριότητα τον πυρήνα της οποίας αποτελεί η συστηματική αξιολόγηση, η οποία είναι μία από τις τέσσερις βασικές μορφές *μέτρησης* (measurement)⁴. Καθώς η μέτρηση είναι ίδιον –δηλαδή ειδοποιό χαρακτηριστικό– της εμπειρικής επιστημονικής έρευνας, η πιστοποίηση της γλωσσομάθειας αντιμετωπίζεται εδώ ως επιστημονική έρευνα. Και κάθε επιστημονική έρευνα κρίνεται από έναν μικρό αριθμό ιδιοτήτων (ή χαρακτηριστικών) που «οφείλει» να έχει, η σημαντικότερη από τις οποίες, σύμφωνα με την καθολική αποδοχή των ειδικών, είναι η εγκυρότητα.

Ο βαθμός εγκυρότητας της ερευνητικής εργασίας που λαμβάνει χώρα στο πλαίσιο της πιστοποίησης της γλωσσομάθειας αποτελεί και το βαθμό εγκυρότητας του κάθε συστήματος πιστοποίησης καθώς και των πιστοποιητικών ή διπλωμάτων ή τίτλων που αυτό χορηγεί.

¹ Με όλο το σεβασμό στην ευαισθησία των αναγνωστριών –που άλλωστε είναι βέβαιο ότι θα είναι περισσότερες από τους αναγνώστες– και στην απαίτησή τους να μην αγνοείται το φύλο τους, θα χρησιμοποιήσω σε όλο το κείμενο το αρσενικό γραμματικό γένος για λόγους καθαρά πρακτικούς: η ταυτόχρονη κάθε φορά αναφορά και στα δύο φύλα κάνει, νομίζω, τα κείμενα κουραστικά.

² Η εγκυρότητα, ως όρος της ελληνικής για να αποδοθεί ο αγγλικός όρος validity, έχει καθιερωθεί εδώ και μερικές δεκαετίες και επομένως θεωρώ ότι δεν υπάρχει λόγος να τον αντικαταστήσουμε με κάποιον άλλο, όπως ο όρος «κύρος» τον οποίο φαίνεται να προτιμά ο Π. Γεωργούσης (1999: 214-262).

³ Αντί για «έλεγχο» της εγκυρότητας θα μπορούσε να μιλήσει κανείς και για «εκτίμηση» ή «αξιολόγηση» (π.χ. Τσοπάνογλου, 2007: 10) της εγκυρότητας. Όπως και να το πει κανείς σωστό είναι. Εδώ θα υιοθετείται, κατά προτίμηση, ο συνηθέστερος, ο πιο καθιερωμένος ή δόκιμος όρος της ελληνικής, επειδή είμαι πεπεισμένος ότι έχει μεγαλύτερη σημασία το σημεινόμενο από το σημαίνον των όρων. Συνεπώς θα δοθεί έμφαση στην όσο γίνεται σαφέστερη διευκρίνιση της κάθε έννοιας αποφεύγοντας τόσο την καταγραφή των διαφόρων εκδοχών του κάθε όρου όσο και τις πολυάριθμες βιβλιογραφικές παραπομπές.

⁴ Τα επιχειρήματα υπέρ αυτής της θέσης παρουσιάζονται στο Τσοπάνογλου (2010²: 34 και 71-78).

Η εξέταση της αξιολόγησης –και ειδικότερα στην εκπαίδευση– σαν να ήταν επιστημονική έρευνα δε σημαίνει πως δεν αναγνωρίζεται εδώ η ύπαρξη και ιδιαιτεροτήτων⁵. Οι ιδιαιτερότητες, ωστόσο, αυτές θεωρώ ότι δεν είναι τέτοιες που να επιβάλουν διαφορετική αντιμετώπιση της αξιολόγησης με στόχο την πιστοποίηση από την αντιμετώπιση της εμπειρικής επιστημονικής έρευνας.

Η εγκυρότητα, με τις διάφορες υποδιαιρέσεις ή διαστάσεις της, και ο έλεγχος της εγκυρότητας είναι έννοιες που δύσκολα μπορεί να εξετάσει κανείς χωριστά, αφού η κατανόηση της μίας συχνά απαιτεί αναφορά στην άλλη. Για παράδειγμα, ο γνωστότερος και επί δεκαετίες οικουμενικά αποδεκτός από τους εφαρμοσμένους γλωσσολόγους ορισμός της γενικής εγκυρότητας, έτσι όπως τον αποτύπωσε ο R. Lado (1961: 321) είναι «Μετράει ένα τεστ αυτό που υποτίθεται ότι μετράει; Αν ναι, τότε είναι έγκυρο»⁶. Είναι προφανές ότι στην προσπάθεια ορισμού της εγκυρότητας ο Lado δίνει τη βασική ερώτηση που θέτει στον εαυτό του όποιος επιχειρεί να ελέγξει την εγκυρότητα. Επομένως ο χωρισμός του κειμένου σε διαφορετικά κεφάλαια για την εγκυρότητα και για τον έλεγχό της έχει σχετική μόνο αξία, στο βαθμό που συχνά το περιεχόμενο του ενός κεφαλαίου θα μπορούσε να ενσωματωθεί στο άλλο.

Η εγκυρότητα και ο έλεγχός της εξετάζονται εδώ από μία οπτική που «παντρεύει» τη διαχρονική και τη συγχρονική μελέτη του θέματος. Δηλαδή αποφεύγεται η παράθεση των απόψεων των ειδικών με χρονική σειρά, που θα έδινε στον αναγνώστη χρήσιμες πληροφορίες για την εξέλιξη ή αλλαγή του θέματος με το πέρασμα του χρόνου⁷, αλλά αποφεύγεται και η έκθεση –και άκριτη αποδοχή– των απόψεων των ειδικών της δεκαετίας που διανύουμε.

Μια τελευταία εισαγωγική διευκρίνιση που χρειάζεται να γίνει είναι πως η εξέταση της εγκυρότητας και του ελέγχου της γίνεται εδώ μέσα από το πρίσμα της αγγλόγλωσσης βιβλιογραφίας. Δηλαδή, στην εξέταση του θέματος σημείο εκκίνησης αποτελεί η ορολογία στην αγγλική γλώσσα για να επιζητηθούν αντίστοιχοι ελληνικοί όροι. Αυτό γίνεται βασικά επειδή οι αγγλοσάξονες έχουν σαφώς πλουσιότερη και καλύτερη βιβλιογραφία στους τομείς της μεθοδολογίας της έρευνας, της *μετρολογίας* (measurement theory) και της *δοκιμασιολογίας* (testing) από εμάς και από άλλους. Αυτό δε σημαίνει ότι στη γαλλόγλωσση⁸, στη γερμανόγλωσση κ.τ.λ. βιβλιογραφία δεν μπορεί να βρει κανείς γνώση

⁵ Τις σημαντικότερες διαφορές, έτσι όπως εμφανίστηκαν στη διεθνή βιβλιογραφία, καταγράφει ο Π. Γεωργούσης (1999: 99-106). Ο συγγραφέας καταλήγει στο συμπέρασμα ότι «η αξιολόγηση δεν είναι έρευνα και οι μέθοδοι της δεύτερης δεν είναι αναγκαίο να υπαγορεύουν τις δραστηριότητες της πρώτης» (1999: 106). Το συμπέρασμα αυτό είναι φυσικά ακριβώς το αντίθετο από τη θέση που υποστηρίζεται εδώ.

⁶ Βλέπε την ακριβή διατύπωση, στην αγγλική, στο επόμενο κεφάλαιο.

⁷ Για διαχρονική εξέταση του θέματος ο αναγνώστης παραπέμπεται στο Chappelle (1999: 254-272), που περιλαμβάνει και αναλυτική σχολιασμένη βιβλιογραφία.

⁸ Σε γαλλική γλώσσα υπάρχουν αρκετά σημαντικά δημοσιεύματα, κυρίως από βέλγους συγγραφείς. Ο γαλλομαθής αναγνώστης θα μπορούσε, μεταξύ άλλων, να συμβουλευτεί τα: De Landsheere 1979: 289-293, De Ketele & Roegiers 1991: 60-64 και 186-206, Van Der Maren 1996: 120-121, 201-202, 361-362, 336-343 και 388-390.

και όρους που θα φώτιζαν πληρέστερα το θέμα. Η εξέταση, ωστόσο, της εγκυρότητας μέσα από το πρίσμα διαφόρων γλωσσών και επιστημονικών παραδόσεων –ιδιαίτερα αν έπαιρνε και διαχρονική μορφή– θα ήταν δύσκολο εγχείρημα, τόσο λόγω του περιορισμένου χώρου όσο και λόγω προσωπικής αδυναμίας, αφού το θέμα δεν εντάσσεται σαφώς στην ειδικότητά μου, που είναι η διδακτική των γλωσσών.

1. Ορισμός της εγκυρότητας

Η εγκυρότητα είναι μία ιδιότητα ή χαρακτηριστικό των επιστημονικών ερευνών που δεν μπορεί να παρατηρηθεί άμεσα. Είναι μια ιδιότητα που ξέρουμε ότι υπάρχει, αφού εμείς τη δημιουργήσαμε, επινοώντας μια λέξη για να κάνουμε αναφορά σε αυτήν, αλλά δύσκολα μπορούμε να δώσουμε έναν καθολικά αποδεκτό ορισμό της.

Κάθε φορά που το αντικείμενο μελέτης είναι μία ιδιότητα μη άμεσα παρατηρήσιμη υπάρχει δυσκολία στον ορισμό της. Για παράδειγμα, μια άλλη τέτοια ιδιότητα είναι η γλωσσομάθεια. Άλλες ακόμη η ευφυΐα ή, καλύτερα, νοημοσύνη, η εξωστρέφεια, η δημιουργικότητα, κ.τ.λ.

Για να γίνει καλύτερα κατανοητή η παραπάνω δήλωση θα πάρουμε τα πράγματα από την αρχή. Κάθε έρευνα έχει ένα ή περισσότερα *υποκείμενα* (subjects), που μας ενδιαφέρουν και που μπορεί να είναι άνθρωποι, πετρώματα, εκπαιδευτικά συστήματα, άλλες έρευνες, κ.τ.λ. Στις εμπειρικές έρευνες δεν ασχολούμαστε με τα υποκείμενα στο σύνολό τους, αλλά εστιάζουμε και μετράμε μία ή περισσότερες ιδιότητες των υποκειμένων.

Όλες οι ιδιότητες των υποκειμένων εντάσσονται σε τρεις μεγάλες κατηγορίες. Υπάρχουν οι ιδιότητες που δεν αλλάζουν από υποκείμενο σε υποκείμενο. Αυτές λέγονται *σταθερές* (constants) και μας ενδιαφέρουν συνήθως λιγότερο από τις υπόλοιπες ή και δε μας ενδιαφέρουν καθόλου. Για παράδειγμα, στην περίπτωση της αξιολόγησης της γλωσσομάθειας με στόχο την πιστοποίηση, σταθερές ιδιότητες των υποκειμένων (δηλαδή εδώ των υποψηφίων ή εξεταζομένων) είναι ότι πρόκειται για ζωντανούς οργανισμούς, ανήκουν όλοι στο ανθρώπινο είδος, κ.τ.λ. Μια ιδιότητα μπορεί, φυσικά, να είναι σταθερά σε μία έρευνα και να μην είναι σταθερά σε μια άλλη.

Οι ιδιότητες που αλλάζουν, δηλαδή που παίρνουν διάφορες τιμές ή υπάρχουν σε διαφορετικό βαθμό σε κάθε υποκείμενο (θεωρώντας και την ανυπαρξία τους, το μηδέν, ως έναν βαθμό), υποδιαιρούνται σε αυτές που είναι άμεσα παρατηρήσιμες και σε αυτές που δεν είναι. Οι πρώτες λέγονται *μεταβλητές* (variables). Μεταβλητές σε μία έρευνα μπορεί να είναι το φύλο, η θερμοκρασία, το εισόδημα, η ηλικία, κ.τ.λ.

Αυτές που δε φαίνονται άμεσα μπορούν να ονομαστούν έννοιες ή δομές ή κατασκευές ή εννοιολογικές δομές. Εδώ θα υιοθετηθεί ο όρος εννοιολογική δομή και για συντομία *δομή* (construct). Στη μεθοδολογία της έρευνας οι ιδιότητες αυτές, στην αγγλική ονομάζονται συχνά και concepts, όρος που δε χρησιμοποιείται καθόλου στη δοκιμασιολογία, όπου

βρίσκει κανείς πάντα τον όρο constructs. Δε θα μπορούμε στον πειρασμό της λεπτομερούς εξέτασης των διαφορών των δύο όρων⁹. Θα εστιάσουμε στο construct, αφού στην προσπάθεια ορισμού της εγκυρότητας ως σφαιρικής ή συνεκτικής έννοιας, διατυπώθηκε εδώ η θέση ότι η εγκυρότητα είναι δομή, δηλαδή σύνθετη και μη άμεσα παρατηρήσιμη ιδιότητα των ερευνών του συγκεκριμένου τύπου που μας απασχολεί εδώ.

Στο λεξικό *γλωσσικής δοκιμασιολογίας* (language testing) των Davies et al. (1999: 31) το construct ορίζεται ως «Η ιδιότητα ή οι ιδιότητες που ένα τεστ σκόπευε να μετρήσει. Μια δομή μπορεί να οριστεί ως μία ικανότητα ή ένα σύνολο ικανοτήτων που αντικατοπτρίζεται στην απόδοση στο τεστ, και για την οποία μπορούμε να κάνουμε συμπερασμούς με βάση την επιτευχθείσα βαθμολογία. Μια δομή ορίζεται γενικώς βάσει μιας θεωρίας· στην περίπτωση της γλωσσομάθειας, βάσει μιας γλωσσολογικής θεωρίας»¹⁰.

Στην έρευνα, κάθε δομή, αφού δεν είναι άμεσα παρατηρήσιμη και μετρήσιμη, πρέπει να υποδιαιρεθεί, ή καλύτερα να «μεταφραστεί», σε μεταβλητές. Πρέπει, δηλαδή, να ορίσουμε από ποιες μεταβλητές αποτελείται ή, καλύτερα, ποιες μεταβλητές αποτελούν εξωτερική εκδήλωση της δομής. Στη διαδικασία αυτή μπορεί να χρειαστεί να περάσει ο ερευνητής από ενδιάμεσα στάδια, δηλαδή να «σπάσει» ή να «μεταφράσει» τη δομή σε μικρότερες ή επιμέρους δομές, πριν φτάσει στο επιθυμητό αποτέλεσμα που είναι ο εντοπισμός του πεπερασμένου αριθμού μεταβλητών που θα μετρηθούν.

Η εγκυρότητα (όπως και η γλωσσομάθεια), ως δομή, λόγω της δυσκολίας να την ορίσουμε, γίνεται καλύτερα κατανοητή αν εξετάσει κανείς πρώτα τις επιμέρους δομές ή και τις μεταβλητές από τις οποίες αποτελείται. Αυτό οδηγεί στον εντοπισμό των ειδών ή τύπων εγκυρότητας. Η αναζήτηση στην αγγλόγλωσση βιβλιογραφία επέτρεψε τον εντοπισμό των παρακάτω 45 τύπων εγκυρότητας:

⁹ Στο παλιό, αλλά εξαιρετικά χρήσιμο ακόμη, λεξικό του De Landsheere, γραμμένο στη γαλλική αλλά με ορισμό και αγγλικών όρων μεθοδολογίας της έρευνας, στο λήμμα concept (1997: 53) διευκρινίζεται ότι «η δομή (ή υποθετική έννοια) διακρίνεται από την έννοια στο βαθμό που δομείται, όχι από την παρατήρηση των ίδιων των αντικειμένων, αλλά από την παρατήρηση εκφάνσεων που αποδίδονται σε ένα αντικείμενο (για παράδειγμα η νοημοσύνη)» [Le construct (ou concept hypothétique) se distingue du concept en ce sens qu'il se construit, non pas à partir de l'observation des objets mêmes, mais de l'observation de manifestations attribuées à un objet (par exemple l'intelligence)]. Αν αποδεχτούμε αυτή τη θέση καθώς και τον ορισμό του concept, που προηγείται του συγκεκριμένου αποσπάσματος, φτάνουμε στο συμπέρασμα ότι το concept είναι μια ιδιότητα μεταξύ του construct και του variable, ως προς το βαθμό σαφήνειας στον ορισμό και ως προς τη δυνατότητα άμεσης παρατήρησης (πρβ. και Quivy & van Campenhoudt 1988: 114-118).

¹⁰ The trait or traits that a test is intended to measure. A construct can be defined as an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores. A construct is generally defined in terms of a theory; In the case of language, a theory of language.

- | | |
|--|-------------------------------------|
| 1) Argument validity | 24) Incremental validity |
| 2) Conceptual validity | 25) Intentional validity |
| 3) Conclusion validity | 26) Internal validity |
| 4) Concrete validity | 27) Intrinsic validity |
| 5) Concurrent validity | 28) Logical validity |
| 6) Congruent validity | 29) Marking validity |
| 7) Consequential validity | 30) Measurement validity |
| 8) Construct validity | 31) Nomological validity |
| 9) Content validity | 32) Observation validity |
| 10) Context validity | 33) Pattern matching validity |
| 11) Convergent validity | 34) Population validity |
| 12) Criterion/-a (-related/-oriented) validity | 35) Predictive validity |
| 13) Deductive validity | 36) Qualitative validity |
| 14) Differential validity | 37) Representation validity |
| 15) Discriminant validity | 38) Scoring validity |
| 16) Divergent validity | 39) Social validity |
| 17) Ecological validity | 40) Statement validity |
| 18) Empirical validity | 41) Statistical conclusion validity |
| 19) Experimental validity | 42) Synthetic validity |
| 20) External validity | 43) Test validity |
| 21) Face validity | 44) Theory-based validity |
| 22) Factorial validity | 45) Translation validity |
| 23) Impact validity | 46) Washback validity |

Θα χάναμε το στόχο μας αν εξετάζαμε εδώ όλους αυτούς τους όρους. Επίσης, κάποιιοι από αυτούς σχετίζονται με την εγκυρότητα στη φιλοσοφία και ειδικότερα στη λογική, που είναι έννοια σχετικά διαφορετική και δε μας ενδιαφέρει. Τέλος, ένας λόγος για τον οποίο δε θα εξετάσουμε κάθε έναν από αυτούς τους όρους είναι το γεγονός ότι κάποιιοι έχουν παρόμοιο αναφερόμενο με άλλους ή είναι υπερώνυμο άλλων. Οι σημαντικοί τύποι εγκυρότητας είναι περίπου 12, οι περισσότεροι από τους οποίους εξετάζονται στο επόμενο κεφάλαιο. Πριν, ωστόσο, γίνει αυτό θα μεταφερθούν εδώ, αυτούσιοι, μερικοί ορισμοί της εγκυρότητας, ως απόδειξη του ισχυρισμού ότι είναι δύσκολο να δώσει κανείς έναν γενικά αποδεκτό ορισμό μιας δομής.

Το 1927 ο Kelly έγραψε: «Το ζήτημα της εγκυρότητας είναι αν ένα τεστ πραγματικά μετράει αυτό που δηλώνει ή φαίνεται πως μετράει»¹¹ (Weir, 2005: 12). Ο ορισμός αυτός υιοθετήθηκε επί δεκαετίες και έδωσε τον, πολύ γνωστό στη σχετική βιβλιογραφία, ορισμό του Lado: «Does a test measure what it is supposed to measure? If it does, it is valid» (Lado, 1961: 321).

¹¹ The problem of validity is that of whether a test really measures what it purports to measure.

Και στους δύο παραπάνω ορισμούς, η εγκυρότητα παρουσιάζεται ως ιδιότητα των τεστ, δηλαδή των οργάνων μέτρησης. Αφήνει έξω τόσο τη χρήση των τεστ όσο και την ερμηνεία των βαθμών που προκύπτουν ή τις συνέπειες που προκαλούν στους εξεταζόμενους, στην εκπαιδευτική διαδικασία και στην κοινωνία γενικότερα. Με άλλα λόγια, αν αποδεχτεί κανείς αυτούς τους ορισμούς θα φτάσει στο συμπέρασμα ότι πολλοί από τους παραπάνω 45 τύπους εγκυρότητας δεν έχουν καμία σχέση με την εγκυρότητα.

Το 1999, από τον Davies και τους συνεργάτες του, η εγκυρότητα ορίζεται ως: «Η ιδιότητα που περισσότερο επηρεάζει την αξία ενός τεστ, ιδιότητα ανώτερη αλλά εξαρτώμενη από την αξιοπιστία. Μια μέτρηση είναι έγκυρη αν κάνει αυτό που υπήρχε πρόθεση να κάνει, δηλαδή να δρα χαρακτηριστικά ως ενδείκτης μιας αφηρημένης έννοιας... την οποία δηλώνει ότι μετράει»¹² (Davies et al., 1999: 221).

Από αυτόν τον ορισμό, που αποτελεί «απόσταγμα» των επικρατούσων απόψεων των δεκαετιών '80 και '90 πρέπει να συγκρατήσουμε το γεγονός ότι μεταθέτει το ενδιαφέρον από το όργανο μέτρησης (test) στη μέτρηση (measure), δηλαδή στη χρήση του οργάνου. Επίσης, διαπιστώνει κανείς ότι κάνει ακόμη πιο σαφή διάκριση μεταξύ εγκυρότητας και αξιοπιστίας (reliability). Πρόκειται για διάκριση που τείνει να εξαφανιστεί στη δεκαετία που διατρέχουμε.

Το 1989 ο Messick και το 1990 ο Bachman, εξετάζοντας την εγκυρότητα, αφενός την όρισαν ως ενιαία και *πολυδιάστατη* (multifaceted) έννοια και αφετέρου έστρεψαν την προσοχή των ειδικών στον τρόπο ερμηνείας των αποτελεσμάτων της μέτρησης και στις συνέπειές της. Αυτά που έγραψαν έγιναν ουσιαστικά κατανοητά και υιοθετήθηκαν δέκα περίπου χρόνια αργότερα, για να δώσουν ορισμούς όπως «Το κύρος ενός τεστ αφορά στην ακρίβεια των συμπερασμάτων ή ερμηνειών, που ένας χρήστης του τεστ καταλήγει από τους βαθμούς των εξεταζομένων στο τεστ»¹³ (Γεωργούσης, 1999: 216), ή «Εγκυρότητα: η καλύτερη διαθέσιμη προσέγγιση στο αληθές ενός δεδομένου ισχυρισμού, ενός λογικού συμπερασμού, ή ενός συμπεράσματος»¹⁴ (Trochim, 2006: κεφ. 1.2.4.).

2. Οι διαστάσεις ή συνισταμένες της εγκυρότητας

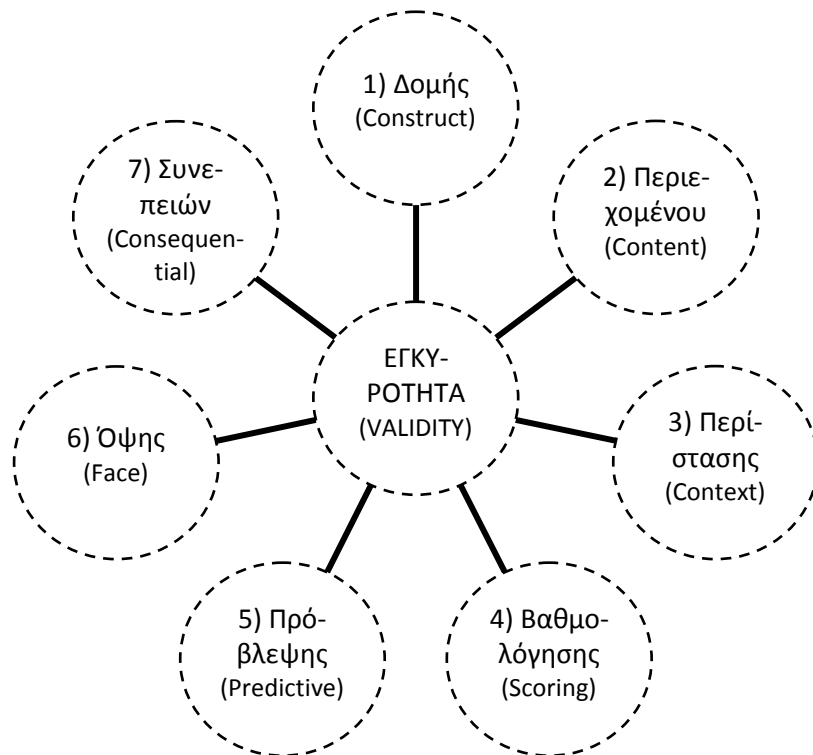
Ακολουθώντας τις απόψεις των Messick και Bachman για την περιεκτικότητα και το πολυδιάστατο της εγκυρότητας, θα εξετάσουμε εδώ μερικούς τύπους εγκυρότητας, εντάσσοντάς τους σε ένα ακτινωτό σχήμα, το κέντρο του οποίου αποτελεί η εγκυρότητα του *court*. Το σχήμα αυτό θα το «διαβάσουμε» με τη φορά των δεικτών ενός ρολογιού, δηλαδή θα ξεκινήσουμε με την εγκυρότητα δομής και θα φτάσουμε στην εγκυρότητα συνεπειών.

¹² The quality which most affects the value of a test, prior to, though dependent on, reliability. A measure is valid if it does what it is intended to do, which is typically to act as an indicator of an abstract concept ... which it claims to measure.

¹³ Μιλώντας για την εγκυρότητα ως χαρακτηριστικό του τεστ, ο Γεωργούσης παραμένει βέβαια, σε κάποιο βαθμό, προσκολλημένος σε θέση προγενέστερη αυτής των Messick και Bachman.

¹⁴ Validity: the best available approximation to the truth of a given proposition, inference, or conclusion.

Με άλλα λόγια, και για να επανέλθουμε στην εξέταση της εγκυρότητας ως δομής, προτείνω εδώ έναν τρόπο απεικόνισης της δομής «εγκυρότητα» καταγράφοντας τις επτά βασικές μεταβλητές (ή απλούστερες δομές: αυτό θα μπορούσε να αποτελέσει αντικείμενο εξέτασης για κάθε μία χωριστά) από τις οποίες εκτιμώ ότι συναποτελείται ή αντιπροσωπεύεται, στην περίπτωση της γλωσσικής πιστοποίησης.



Σχήμα 1: Απεικόνιση των διαστάσεων της εγκυρότητας

Η επιλογή των συγκεκριμένων διαστάσεων της εγκυρότητας και η αποσιώπηση άλλων έγινε με κριτήριο τη σημασία τους για το συγκεκριμένο τύπο ερευνητικής δραστηριότητας. Είναι προφανές ότι, ανάλογα με τον τύπο έρευνας, η κάθε διάσταση της εγκυρότητας αποκτά διαφορετικό βάρος, ή και καμιά φορά χάνει κάθε νόημα. Για παράδειγμα, σε μία έρευνα του τύπου *μελέτη περίπτωσης* (case study), η *εξωτερική εγκυρότητα* (external validity), που είναι ο βαθμός στον οποίο τα αποτελέσματα μπορούν να γενικευτούν, δεδομένης της ποιότητας και της ποσότητας του *δείγματος* (sample) υποκειμένων που παρατηρήθηκαν¹⁵, δεν έχει σημασία. Αντίθετα, σε μία έρευνα επί δείγματος, του τύπου survey, η εξωτερική εγκυρότητα ή *εγκυρότητα πληθυσμού* (population validity) είναι προφανώς κεφαλαιώδους σημασίας.

¹⁵ Με τον όρο εξωτερική εγκυρότητα γίνεται εδώ αναφορά σε κάτι διαφορετικό από αυτό που ορίζεται στο λεξικό των Davies et al. Στο λεξικό η εξωτερική εγκυρότητα ταυτίζεται με την *εγκυρότητα κριτηρίου* (criterion-related validity). Συνοπτικότερος και πλησιέστερος στη γενική εκδοχή του όρου στη μεθοδολογία της έρευνας είναι ο ορισμός που έχει καταχωρηθεί στη Wikipedia: «The issue of External validity concerns the question to what extent one may safely generalize the (internally valid) causal inference (a) from the sample studied to the defined target population and (b) to other populations (i.e. across time and space)» (Wikipedia, 06/08/08).

2.1 Εγκυρότητα δομής

Η πρώτη διάσταση της εγκυρότητας είναι η *εγκυρότητα δομής* (construct validity). Με τον όρο αυτό συχνά, τα τελευταία χρόνια, γίνεται αναφορά στην εγκυρότητα στο σύνολό της, αφού στο τέλος της δεκαετίας του '80 υπήρξε «αναβάθμιση» της εγκυρότητας δομής, που από τύπος εγκυρότητας έγινε, για ορισμένους ειδικούς, υπερώνυμο πολλών τύπων και σχεδόν αντικατέστησε την εγκυρότητα (Anastasi 1986 και 1988). Εδώ θα υιοθετηθεί προγενέστερη εκδοχή του όρου και λέγοντας εγκυρότητα δομής θα κάνουμε αναφορά στο βαθμό που «...ένα τεστ μπορεί να ερμηνευτεί ως μέτρο κάποιου χαρακτηριστικού ή κάποιας ιδιότητας που δεν 'έχει οριστεί σε επίπεδο μεταβλητών'»¹⁶ (Cronbach & Meehl, 1955: 282).

Παρόμοια και συμπληρωματική εκδοχή του όρου βρίσκει κανείς και στο λεξικό των Davies et al.: «Η εγκυρότητα δομής ενός γλωσσικού τεστ είναι μια ένδειξη του κατά πόσο αυτό αντιπροσωπεύει μία υφιστάμενη θεωρία εκμάθησης της γλώσσας»¹⁷ (1999: 33).

Με βάση την παραπάνω αποδοχή, και παίρνοντας ως πεδίο εφαρμογής την πιστοποίηση της γλωσσομάθειας, μπορεί να ισχυριστεί κανείς ότι η εγκυρότητα δομής εξαρτάται από τη συνέπεια και το βάσιμο μετατροπής της δομής «γλωσσομάθεια» ή «γνώση ξένων γλωσσών»¹⁸ –που είναι δοσμένη, δηλαδή αποφασισμένη από τον φορέα πιστοποίησης με «πολιτική» βούληση– σε ένα σύνολο από προδιαγραφές στις οποίες τα τεστ, τα όργανα μέτρησης που κατασκευάζονται για κάθε εξεταστική περίοδο, οφείλουν να υπακούουν.

Οι προδιαγραφές των οργάνων μέτρησης ενός συστήματος πιστοποίησης γλωσσομάθειας, για κάθε επίπεδο, πρέπει να ορίζουν τόσο τις επιμέρους *ικανότητες* (competences) που θα ελέγχονται μέσω κάθε *δοκιμασίας* (exercise/activity), όσο και του τρόπου με τον οποίο θα γίνεται αυτός ο έλεγχος, δηλαδή της τυπολογίας δοκιμασιών, του χρόνου που θα διαθέτουν οι εξεταζόμενοι, της μορφής των *οδηγιών* (instructions) κάθε δοκιμασίας, κ.τ.λ. Με άλλα λόγια, δεν αρκεί η μετατροπή της δομής σε μεταβλητές, διότι χρειάζεται παράλληλα να ορίζεται με ακρίβεια ο τρόπος, η μέθοδος μέτρησης των μεταβλητών, αφού μια ακατάλληλη μέθοδος μέτρησης μπορεί να αφαιρέσει βαθμούς από την εγκυρότητα δομής επιτρέποντας σε άλλες δομές (όπως η νοημοσύνη, η εμπειρία στην εφαρμογή στρατηγικών αντιμετώπισης ενός τύπου δοκιμασίας, κ.τ.λ.) να επηρεάσουν τη βαθμολογία που θα προκύψει.

Είναι αυτονόητο ότι πριν ακόμη από τη σύνταξη των προδιαγραφών, δηλαδή πριν από τη «μετάφραση» της δομής σε μεταβλητές και σε μέθοδο μέτρησής τους, οι ειδικοί που εργάζονται για το σύστημα πρέπει να επιλέξουν την εκδοχή της «γλωσσομάθεια» που θα

¹⁶ ...a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined."

¹⁷ The construct validity of a language test is an indication of how representative it is of an underlying theory of language learning.

¹⁸ Αξίζει να παρατηρήσει κανείς, υπό μορφή παραδείγματος, ότι το τμήμα του Υπουργείου Εθνικής Παιδείας και Θρησκευμάτων που έχει την ευθύνη λειτουργίας του Κρατικού Πιστοποιητικού Γλωσσομάθειας, ως συστήματος πιστοποίησης, ονομάζεται «Διεύθυνση Πιστοποίησης της Γνώσης Ξένων Γλωσσών».

υιοθετήσουν. Με άλλα λόγια πρέπει να αποφασίσουν ποια/ποιες από τις θεωρίες ανάλυσης της γλώσσας, της γλωσσικής επικοινωνίας και της γλωσσικής εκμάθησης αποδέχονται. Αυτός ο προσδιορισμός της δομής είναι τελείως απαραίτητος και αποτελεί στη συνέχεια το γνώμονα των ειδικών που αναλαμβάνουν το έργο σύνταξης των προδιαγραφών του συστήματος.

Ο προσδιορισμός της δομής πριν από την ανάλυσή της ήταν πριν από μερικές δεκαετίες μία επίπονη εργασία για ένα σοβαρό σύστημα πιστοποίησης. Τις τρεις τελευταίες δεκαετίες η εργασία αυτή διευκολύνεται ιδιαίτερα από την ύπαρξη προτάσεων που ετοιμάζουν φορείς διεθνούς εμβέλειας, ο σημαντικότερος από τους οποίους είναι το Συμβούλιο Πολιτισμικής Συνεργασίας (Council for Cultural Co-operation) του Συμβουλίου της Ευρώπης. Τα δημοσιεύματα του Συμβουλίου, και ιδιαίτερα αυτό που είναι γνωστό ως *Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς για τις Γλώσσες* (ΚΕΠΑ)¹⁹ διευκολύνουν ιδιαίτερα την εργασία των συστημάτων πιστοποίησης. Μπορεί οι ειδικοί κάποιου συστήματος να διαφωνούν και να διαφοροποιηθούν ως προς κάποια σημεία. Διαθέτουν, ωστόσο, μια βάση, ένα επαρκώς ομοιογενές θεωρητικό υπόβαθρο από το οποίο να ξεκινήσουν.

Συμπερασματικά, έτσι όπως βλέπουμε εδώ την εγκυρότητα δομής, για να την εξασφαλίσει κανείς πρέπει να φροντίσει να υπάρχει θετική, ή πάντως ικανοποιητική, απάντηση σε σειρά ερωτημάτων, τα σημαντικότερα από τα οποία είναι τα εξής:

- 1) Τι κρύβεται κάτω από τη δομή που το σύστημα πιστοποίησης (θα) μετράει; Δηλαδή ποιο θεωρητικό πλαίσιο (ή θεωρητική βάση ή θεωρία) υιοθετείται από το σύστημα;
- 2) Είναι το θεωρητικό πλαίσιο επαρκώς ομοιογενές; Μήπως, δηλαδή, το θεωρητικό πλαίσιο περιλαμβάνει επιμέρους θεωρίες που δεν ταιριάζουν μεταξύ τους, δε συνάδουν;
- 3) Ποιες τροποποιήσεις στο θεωρητικό πλαίσιο επιβάλλονται από τις τοπικές συνθήκες (κοινωνικές, πολιτισμικές, οικονομικές, κ.τ.λ.) και τις ιδιαιτερότητες των χρηστών (γνωστικό και μαθησιακό προφίλ λόγω εκπαιδευτικού συστήματος κ.τ.λ.)²⁰;

¹⁹ Πρόκειται για κείμενο που έχει θέση επίσημου εγγράφου, δηλαδή την ευθύνη για το περιεχόμενό του δεν την έχουν μόνο οι συγγραφείς του, αλλά και ο φορέας από τον οποίο εκπορεύεται, δηλαδή το Συμβούλιο της Ευρώπης, αρχικά, και στη συνέχεια και η Ευρωπαϊκή Ένωση που το υιοθέτησε. Η πρώτη έντυπη δημοσίευση του ΚΕΠΑ έγινε το 2001 με τον τίτλο *Common European Framework of Reference for Languages: Learning, teaching, assessment*, αλλά ήδη από το 1995 υπήρχαν στον ιστότοπο του Συμβουλίου της Ευρώπης προγενέστερες εκδοχές του κειμένου στην αγγλική και στη γαλλική γλώσσα. Σήμερα (Νοέμβρης 2011) ο αναγνώστης μπορεί να βρει την αγγλική και γαλλική τελευταία έκδοση στις διευθύνσεις που καταγράφονται στο τέλος αυτού του κειμένου (βλ. βιβλιογραφικές παραπομπές).

²⁰ Αυτό το ερώτημα σχετίζεται σε κάποιο βαθμό με μία διάσταση της εγκυρότητας που είναι τελευταία γνωστή με τον όρο *περιβαλλοντική εγκυρότητα* (ecological validity). Η περιβαλλοντική εγκυρότητα στη Wikipedia (06/08/08) ορίζεται ως εξής: «Ecological validity is whether the results can be applied to real life situations». Ο ορισμός αυτός φαίνεται να ταιριάζει ακόμη και στην εξωτερική αξιολόγηση. Εδώ θα θεωρήσουμε ότι οι δύο διαστάσεις διαφοροποιούνται από το γεγονός ότι η περιβαλλοντική εγκυρότητα λαμβάνει υπόψη τοπικές ιδιαιτερότητες. Δηλαδή, στην περίπτωση της πιστοποίησης, μήπως η γλωσσομάθεια και ειδικότερα, για παράδειγμα, η «αγγλομάθεια» στον ελληνικό χώρο, έχει κάποια χαρακτηριστικά που είναι διαφορετικά από αυτά που έχει, π.χ., στην Ινδία;

- 4) Με βάση το θεωρητικό πλαίσιο, ποιες είναι οι μεταβλητές που το σύστημα θεωρεί ότι αποτελούν εξωτερική εκδήλωση της δομής ή συσχετίζονται με αυτήν;
- 5) Και με ποιο τρόπο (θα) γίνεται η μέτρηση των μεταβλητών, έτσι ώστε η μορφή του οργάνου μέτρησης και η χρήση του να μην επιτρέπουν ουσιαστική²¹ παρείσφρηση άλλων δομών ή μεταβλητών στη διαμόρφωση της βαθμολογίας, δηλαδή του αποτελέσματος ή της ένδειξης;
- 6) Τέλος, μήπως υπάρχουν πτυχές ή διαστάσεις ή μεταβλητές της δομής που δεν έχουν επαρκή αντιπροσώπευση στις προδιαγραφές²² και άλλες που υπερτονίζονται;

Οι μέθοδοι που μπορούν να εξασφαλίσουν απάντηση στα παραπάνω ερωτήματα και το είδος ενδείξεων και αποδείξεων που ένα σύστημα πιστοποίησης οφείλει να παρέχει σε όποιον θα ήθελε να το αξιολογήσει, εξετάζονται στο κεφάλαιο 3.

2.2 Εγκυρότητα περιεχομένου

Οι προδιαγραφές, δηλαδή τα χαρακτηριστικά των οργάνων μέτρησης και της χρήσης τους – που κάθε σύστημα γλωσσομάθειας οφείλει να διαθέτει και να θέτει στη διάθεση του κοινού– θα χαρακτηριστούν εδώ «περιεχόμενο». Η *εγκυρότητα περιεχομένου* (content validity) δεν είναι τίποτε άλλο από το βαθμό στον οποίο το περιεχόμενο «αντιπροσωπείται» στις δοκιμασίες ενός συγκεκριμένου οργάνου μέτρησης ή σε όλα τα όργανα ενός συστήματος πιστοποίησης.

Στη συστηματική εκπαιδευτική αξιολόγηση γενικότερα, τη θέση των προδιαγραφών παίρνει το πρόγραμμα σπουδών. Έτσι, για παράδειγμα, στις εξετάσεις για επιλογή αυτών που θα εγγραφούν στα ιδρύματα τριτοβάθμιας εκπαίδευσης της χώρας, τα Αναλυτικά Προγράμματα Σπουδών κάθε μαθήματος αποτελούν το σύνολο των περιεχομένων, του οποίου συνόλου αντιπροσωπευτικό δείγμα πρέπει να εμφανίζεται στα θέματα εξέτασης κάθε εξεταστικής περιόδου.

Αυτή η διάσταση εγκυρότητας είναι ίσως η ευκολότερα οριζόμενη, και ο ορισμός της είναι γενικά αποδεκτός, με ελάχιστες εξαιρέσεις.

Όπως είναι προφανές, σε ένα σοβαρό σύστημα πιστοποίησης, είναι αδύνατο κάθε όργανο μέτρησης που ετοιμάζεται για μια εξεταστική περίοδο να περιλαμβάνει και να ελέγχει όλα τα προδιαγεγραμμένα περιεχόμενα²³. Με άλλα λόγια, υποχρεωτικά το περιεχόμενο του κάθε οργάνου αποτελεί δείγμα του περιεχομένου της εξέτασης, δηλαδή των

²¹ Η παρείσφρηση άλλων δομών ή/και μεταβλητών στη διαμόρφωση του αποτελέσματος της μέτρησης είναι πρακτικά αδύνατο να αποφευχθεί πλήρως. Η προσπάθεια πρέπει να στοχεύει στην ελαχιστοποίηση αυτής της παρείσφρησης.

²² Δηλαδή, προσπαθούμε να αποφύγουμε την *υπο-αντιπροσώπευση της δομής* (construct under-representation) (πρβ. Messick, 1989).

²³ Φυσικά υπάρχουν και συστήματα πιστοποίησης που δε διαθέτουν προδιαγραφές και όπου σε κάθε όργανο μέτρησης εξετάζονται τα ίδια ακριβώς πράγματα στην ίδια σειρά. Σε τέτοια περίπτωση δεν έχει νόημα ο έλεγχος της εγκυρότητας περιεχομένου, αφού το κάθε τεστ περιλαμβάνει δυνητικά το σύνολο των περιεχομένων, δηλαδή τον *πληθυσμό* (population), και όχι δείγμα, του οποίου την αντιπροσωπευτικότητα θα μπορούσε να ελέγξει κανείς.

προδιαγραφών. Το ζητούμενο είναι το δείγμα αυτό να μην είναι *συμπτωματικό* (casual) ή/και *μεροληπτικό* (biased), αλλά *τυχαίο* (random) ή *αντιπροσωπευτικό* (representative). Οι μέθοδοι *δειγματοληψίας* (sampling), που βρίσκει κανείς σε όλα σχεδόν τα συγγράμματα μεθοδολογίας της έρευνας και στατιστικής²⁴, μπορούν να αποτελέσουν φυσικά χρήσιμο οδηγό για όποιον ασχοληθεί με την εγκυρότητα περιεχομένου ενός συστήματος πιστοποίησης.

Θεωρώ πως η εγκυρότητα περιεχομένου σε αυτό το είδος έρευνας είναι το αντίστοιχο της εξωτερικής εγκυρότητας των δειγματοληπτικών ερευνών τύπου survey. Στις έρευνες τύπου survey το ερώτημα που τίθεται είναι αν μπορεί κανείς να γενικεύσει τα συμπεράσματα στα οποία κατέληξε, δεδομένης της ποσότητας και ποιότητας του δείγματος. Στις έρευνες *αξιολόγησης* (της γλωσσικής) *επάρκειας* (proficiency testing) με στόχο την πιστοποίηση, το ερώτημα που τίθεται είναι ακριβώς το ίδιο. Το μόνο που αλλάζει είναι ότι στην πρώτη περίπτωση η *δειγματοληπτική μονάδα* (sampling unit) είναι το μεμονωμένο υποκείμενο (συνήθως άνθρωπος) που πήραμε από τον πληθυσμό και το περιλάβαμε στο δείγμα, ενώ στη δεύτερη περίπτωση η δειγματοληπτική μονάδα είναι η κάθε δοκιμασία που πήραμε από το σύνολο των δυνατών δοκιμασιών, βάσει των προδιαγραφών, και το περιλάβαμε σε ένα τεστ.

Ταυτόχρονα, όμως, η εγκυρότητα περιεχομένου έχει και κάποια χαρακτηριστικά της *εσωτερικής εγκυρότητας* (internal validity). Εσωτερική εγκυρότητα χαρακτηρίζεται στη μεθοδολογία της έρευνας ο βαθμός στον οποίο τα συμπεράσματα για τη *σχέση αιτίου αιτιατού* (causal inference) είναι πιθανό να είναι αληθή, δεδομένων των οργάνων μέτρησης που χρησιμοποιήθηκαν και του γενικού ερευνητικού σχεδιασμού. Αν κατά τη μέτρηση της *ανεξάρτητης μεταβλητής* (independent variable), δηλαδή του αιτίου, ή της εξαρτημένης μεταβλητής (dependent variable), δηλαδή του αιτιατού ή της συνέπειας, κάναμε το λάθος να περιλάβουμε στο όργανο μέτρησης ερωτήματα που δε σχετίζονται με τις συγκεκριμένες δύο μεταβλητές, τότε λέμε ότι η έρευνα έχει μειωμένη εσωτερική εγκυρότητα. Αντιστοίχως, αν το περιεχόμενο του γλωσσικού τεστ περιλαμβάνει ερωτήματα που δεν υπήρχαν στις προδιαγραφές, επειδή δεν αφορούν κάποια από τις μεταβλητές που συναποτελούν τη δομή ή συσχετίζονται με αυτήν, τότε τα αποτελέσματα της εξέτασης δε θα απεικονίζουν αρκετά πιστά τη δομή, δηλαδή τη γλωσσομάθεια, που ενέχει θέση αιτίου, αν η γλωσσική παραγωγή ή αντίδραση του εξεταζόμενου στην εξέταση θεωρηθεί αιτιατό. Το έκτο από τα παρακάτω ερωτήματα για τον έλεγχο της εγκυρότητας περιεχομένου προκύπτει από αυτήν ακριβώς την μερική επικάλυψη της εγκυρότητας περιεχομένου και της εσωτερικής εγκυρότητας.

Με την ευκαιρία, πρέπει να τονιστεί ότι όλες οι διαστάσεις της εγκυρότητας έχουν στην ουσία πάντα κάποια σχέση μεταξύ τους και συχνά και κάποιο βαθμό επικάλυψης. Αυτό

²⁴ Βλέπε, για παράδειγμα, Blalock 1984: 683-711, Langouet & Porlier 1991: 26-27, 74-77, Τσάντας κ.ά. 1999: 18-24, Verma & Mallick 2004: 314-322.

είναι φυσικό και επιστημονικά αποδεκτό, αφού τα είδη εγκυρότητας δίνουν μία (ή περισσότερες) *τυπολογία* (typology) και όχι μία *ταξινόμια* (taxonomy).

Για να επιστρέψουμε στο κύριο αντικείμενο αυτής της ενότητας του κειμένου, εάν ένα όργανο μέτρησης που ετοιμάστηκε για μια εξεταστική περίοδο διαθέτει εγκυρότητα περιεχομένου, αυτό δε σημαίνει ότι και η συνολική ερευνητική δραστηριότητα που αναπτύσσεται από το σύστημα πιστοποίησης της γλωσσομάθειας διαθέτει εγκυρότητα περιεχομένου. Είναι, προφανώς, απαραίτητο κάθε νέο όργανο μέτρησης που ετοιμάζεται να αποτελεί *διαφορετικό* δείγμα, εξίσου αντιπροσωπευτικό. Με άλλα λόγια, η αντιπροσωπευτικότητα είναι μια ιδιότητα που πρέπει να ελέγχεται και σε βάθος χρόνου.

Συμπερασματικά, τα ερωτήματα που πρέπει να απαντηθούν, αναφορικά με την εγκυρότητα περιεχομένου, είναι:

- 1) Σε ποιο βαθμό οι ελεγχόμενες επιμέρους δεξιότητες ή μεταβλητές σε ένα τεστ αντιπροσωπεύουν το σύνολο των μεταβλητών που υπάρχουν στις προδιαγραφές, έτσι ώστε να μπορεί να ισχυριστεί κανείς ότι, αν ήταν πρακτικά εφικτό το τεστ να ελέγξει/μετρήσει όλες τις μεταβλητές, το αποτέλεσμα της μέτρησης δε θα παρουσίαζε ουσιαστική διαφοροποίηση;
- 2) Και –αντίστοιχα με το παραπάνω ερώτημα– σε ποιο βαθμό τα θέματα συζήτησης ή πεδία αναφοράς στο τεστ αντιπροσωπεύουν το σύνολο των πεδίων αναφοράς για τα οποία περιμένουμε να μπορεί να καταλάβει ή να μιλήσει/γράψει κανείς όταν διαθέτει το/τα συγκεκριμένο/-α επίπεδο/-α γλωσσομάθειας για το/τα οποίο/οποία έγινε το τεστ;
- 3) Επίσης –αντίστοιχα– πόση αντιπροσωπευτικότητα υπάρχει στους τύπους δοκιμασιών που αξιοποιήθηκαν στο συγκεκριμένο τεστ σε σχέση με την τυπολογία δοκιμασιών που ορίζεται στις προδιαγραφές;
- 4) Μήπως υπάρχουν στοιχεία του περιεχομένου των προδιαγραφών που δεν εμφανίζονται σε κανένα από τα όργανα μέτρησης του συστήματος πιστοποίησης που χρησιμοποιήθηκαν τις τελευταίες, για παράδειγμα, δέκα εξεταστικές περιόδους; Αν ναι, είναι δυνατό να δικαιολογηθεί αυτό με κάποιον τρόπο;
- 5) Μήπως υπάρχουν στοιχεία που εμφανίζονται κάθε φορά; Και πόσο δικαιολογημένο είναι αυτό; Είναι, τουλάχιστον, στοιχεία που έχουν ιδιαίτερη σημασία, μεγάλο ειδικό βάρος, για τη δομή που μετριέται μέσω των τεστ, δηλαδή για τη γλωσσομάθεια;
- 6) Μήπως σε κάποιο από τα τεστ που χρησιμοποίησε το σύστημα υπήρχαν δοκιμασίες που μετρούσαν κάποια μεταβλητή, η οποία δεν περιλαμβανόταν καθόλου στις προδιαγραφές;

Η παραπάνω λίστα ερωτημάτων μπορεί να επιμηκυνθεί ανάλογα με τον πλούτο ή την έκταση των προδιαγραφών που το σύστημα πιστοποίησης έχει ετοιμάσει.

2.3 Εγκυρότητα περίπτωσης

Οι πρώτες δύο διαστάσεις της εγκυρότητας έχουν εδώ πιο περιορισμένο αναφερόμενο από αυτό που βρίσκει κανείς στην αγγλόγλωσση βιβλιογραφία, αφού όπως έχει ήδη δηλωθεί, η εγκυρότητα δομής, κατά την άποψη πολλών ειδικών, περιλαμβάνει όλα τα είδη εγκυρότητας. Η εγκυρότητα περιεχομένου, επίσης, την οποία ο Weir μετονομάζει σε context validity (2005: 19), περιλαμβάνει συχνά, εκτός της αντιπροσωπευτικότητας, και την ιδιότητα της καταλληλότητας του τεστ για το κοινό για το οποίο σχεδιάστηκε.

Αξιοποιώντας τις αναλυτικά διατυπωμένες απόψεις του Weir (2005: 56-83) για τον έλεγχο της εγκυρότητας περιεχομένου, προτείνω τη χρήση του όρου *εγκυρότητα περίπτωσης*, με τον οποίο θα γίνεται εδώ αναφορά στο βαθμό που τα επιμέρους χαρακτηριστικά ενός τεστ και οι συνθήκες διεξαγωγής του ταιριάζουν στα χαρακτηριστικά των εξεταζομένων. Όπως είναι προφανές, εάν ένα όργανο μέτρησης έχει χαρακτηριστικά και χρησιμοποιείται με τρόπο που να προκαλούν αρνητική αντίδραση –ή την αδράνεια– των υποκειμένων της έρευνας, δηλαδή των εξεταζομένων, τα αποτελέσματα της όλης διαδικασίας δε θα απεικονίζουν το βαθμό γλωσσομάθειάς τους και η όποια ερμηνεία τους θα είναι επισφαλής.

Σύμφωνα με την οπτική εξέτασης της εγκυρότητας που αναπτύσσεται εδώ, η εγκυρότητα δομής και η εγκυρότητα περιεχομένου είναι ιδιότητες που υπάρχουν και που μπορούν να ελεγχθούν πριν από την υλοποίηση της μέτρησης, δηλαδή πριν από τη διεξαγωγή εξετάσεων. Τα στοιχεία ή αντικείμενα, πράγματι, στα οποία μπορεί να στηριχτεί κανείς για να ελέγξει τα πρώτα δύο είδη εγκυρότητας (σαφής προσδιορισμός της δομής, μετάφρασή της σε μεταβλητές βάσει συγκεκριμένης θεωρίας, μετουσίωσή της σε όργανο μέτρησης που να περιλαμβάνει αντιπροσωπευτικό δείγμα των μεταβλητών και των άλλων περιεχομένων των προδιαγραφών) δεν απαιτούν δεδομένα από την ίδια την εξέταση. Η τρίτη, αντίθετα, αυτή διάσταση της εγκυρότητας, όπως και η τέταρτη, ελέγχονται με βάση στοιχεία που προκύπτουν κατά τη διάρκεια ή αμέσως μετά από τη διεξαγωγή της εξέτασης.

Με την ευκαιρία της συσχέτισης των διαστάσεων της εγκυρότητας με το χρόνο συλλογής των δεδομένων για τον έλεγχό τους, είναι σκόπιμο να αναφερθούμε και σε όλες τις άλλες. Πρέπει να σημειώσουμε ότι, ενώ οι δύο πρώτες επιτρέπουν έλεγχο πριν τη διεξαγωγή της μέτρησης, οι διαστάσεις 3 και 4 απαιτούν τη διεξαγωγή της μέτρησης και οι διαστάσεις 5, 6 και 7 χρειάζονται συλλογή πρόσθετων δεδομένων μετά τη διεξαγωγή της μέτρησης, τα οποία αντιπαραβάλλονται ή συσχετίζονται τόσο με ό,τι υπάρχει πριν από τη μέτρηση όσο και, κυρίως, με τα ίδια αποτελέσματα της μέτρησης. Το θέμα αυτό θα γίνει σαφέστερο στο κεφάλαιο 3.

Μερικά σημαντικά ερωτήματα, η απάντηση των οποίων δίνει το μέτρο της εγκυρότητας περίπτωσης, είναι:

- 1) Πόσο εύκολα κατανοητές και αποδεκτές από τους εξεταζόμενους είναι οι οδηγίες εκτέλεσης των δοκιμασιών; Δηλαδή καταλαβαίνουν αμέσως τι πρέπει να κάνουν και βρίσκουν λογικό και σκόπιμο ό,τι πρέπει να κάνουν;
- 2) Πόσο φιλική προς τους συγκεκριμένους εξεταζόμενους, δηλαδή προς το κοινό για το οποίο κατασκευάστηκε το τεστ, είναι η σελιδοποίηση (μέγεθος χαρακτήρων, χρησιμοποιούμενες γραμματοσειρές, αριθμός ερωτημάτων σε κάθε σελίδα, στοιχεία διάκρισης της αρχής και του τέλους κάθε δοκιμασίας, κ.τ.λ.);
- 3) Πόσο ενδιαφέρουσες βρίσκουν οι εξεταζόμενοι τις δοκιμασίες, δεδομένου του τύπου και του αριθμού τους; Υπάρχουν δοκιμασίες που προκάλεσαν κάποιο κίνητρο και μεγαλύτερο ενδιαφέρον από άλλες;
- 4) Η σειρά εμφάνισης των δοκιμασιών συνέβαλε στην ομαλή διεξαγωγή της εξέτασης ή μήπως, για το συγκεκριμένο κοινό εξεταζομένων, θα έπρεπε να είναι διαφορετική;
- 5) Υπάρχουν ενδείξεις ανεπάρκειας ή υπερεπάρκειας του χρόνου για τη διεξαγωγή συγκεκριμένων δοκιμασιών ή τμημάτων/ενοτήτων του τεστ; Μήπως, δηλαδή, για το συγκεκριμένο κοινό (ηλικία, σχολικές συνήθειες, επίπεδο εκπαίδευσης, κ.τ.λ.) ο χρόνος δε φτάνει (προκαλώντας άγχος) ή είναι χωρίς λόγο πλεονάζων;
- 6) Μήπως κάποιο χαρακτηριστικό του τεστ και της περίπτωσης (έκταση κειμένων, θεματολογία, στάση και συμπεριφορά εξεταστών του προφορικού λόγου, χαρακτήρας φωτογραφιών που χρησιμοποιούνται ως ερέθισμα, κ.τ.λ.) προκαλεί αντιδράσεις λόγω του πολιτισμικού υπόβαθρου του κοινού;

Όπως είναι προφανές η κατάσταση αυτή θα μπορούσε να επεκταθεί, αφού ήδη το τελευταίο ερώτημα είναι συνεκτικό και θα μπορούσε να υποδιαιρευθεί σε τέσσερα τουλάχιστον ερωτήματα.

2.4 Εγκυρότητα βαθμολόγησης

Με τον όρο αυτό, που προτείνεται ως αντίστοιχος του *scoring validity* του Weir (2005: 177-206), θα κάνουμε αναφορά στο βαθμό που η βαθμολογία που προκύπτει από μία μέτρηση απεικονίζει ή αποδίδει τη γλωσσομάθεια κάθε εξεταζόμενου και αυτή η βαθμολογία θα ήταν (σχεδόν) η ίδια αν η μέτρηση επαναλαμβανόταν χωρίς να μεσολαβήσει μάθηση.

Όπως είναι φανερό, η εγκυρότητα βαθμολόγησης δεν είναι άλλο από αυτό που συνήθως καλείται αξιοπιστία του τεστ.

Έναν άλλο ορισμό της εγκυρότητας βαθμολόγησης που θα μπορούσε να δώσει κανείς, τολμώντας να διαφοροποιηθεί από τα καθιερωμένα, είναι ο εξής: εγκυρότητα βαθμολόγησης –στην αξιολόγηση/μέτρηση της γλωσσομάθειας με στόχο την πιστοποίηση– είναι η συνέπεια με την οποία γίνεται μετατροπή των ποιοτικών δεδομένων σε ποσοτικά. Για να γίνει αυτό κατανοητό θα ξεκινήσουμε από προγενέστερη στιγμή της ερευνητικής δραστηριότητας.

Στις επιστημονικές έρευνες, όταν γίνεται μία μέτρηση, άλλοτε συλλέγουμε ποιοτική και άλλοτε ποσοτική πληροφορία. Ποιοτική πληροφορία έχουμε όταν η μέτρηση γίνεται χρησιμοποιώντας *ονοματική κλίμακα* (nominal scale). Πάνω σε αυτή την πληροφορία δεν μπορούμε να κάνουμε στατιστική ανάλυση, ούτε καν τις βασικές αριθμητικές πράξεις. Ποσοτική, από την αρχή, πληροφορία παίρνουμε όταν χρησιμοποιούμε, κατά τη μέτρηση, *(ισο)διαστημική κλίμακα* (interval scale) ή *αναλογική κλίμακα* (ratio scale). Πάνω σε αυτή την πληροφορία μπορούμε αμέσως να κάνουμε αριθμητικές πράξεις.

Ένα τέταρτο είδος κλίμακας που έχει επινοηθεί λέγεται *τακτική κλίμακα* (ordinal scale) και, κατά την άποψη των περισσότερων ειδικών, δίνει επίσης πληροφορία ποιοτικής φύσης. Εδώ, ωστόσο, υπάρχει και αντίλογος. Δε θα ασχοληθούμε με αυτό γιατί δεν είναι σκόπιμο να χαθούμε σε επιστημολογικής φύσης θέματα. Εδώ θα θεωρήσουμε –συμβατικά ή για ευκολία μας, κάνοντας απλοποίηση– ότι η πληροφορία είναι ποσοτική, στο βαθμό που εκφράζεται με ένα νούμερο που μας επιτρέπει να κάνουμε αριθμητικές πράξεις, έστω και καταχρηστικά. Επομένως, θα θεωρήσουμε ότι η διάκριση ποιοτικής και ποσοτικής πληροφορίας θα γίνεται εδώ με βάση τη δυνατότητα εκτέλεσης αριθμητικών πράξεων.

Στην περίπτωση της μέτρησης της γλωσσομάθειας όλες οι πληροφορίες που παίρνουμε κατά την εξέταση από κάθε εξεταζόμενο είναι καταρχήν ποιοτικές. Δηλαδή ο λόγος τον οποίο παράγει ο εξεταζόμενος (είτε γραπτός είτε προφορικός) έχει κυρίως ποιοτικά χαρακτηριστικά και πάντως σε αυτά εστιάζουμε. Αλλά και όταν δεν παράγει λόγο αλλά απλά επιλέγει την απάντηση «b» από ένα σύνολο τεσσάρων δυνατών επιλογών («a», «b», «c», «d»), αυτό που μας δίνει είναι μία ποιοτική πληροφορία πάνω στην οποία δεν μπορούμε να κάνουμε καμία αριθμητική πράξη.

Η βαθμολόγηση είναι η μετατροπή της ποιοτικής πληροφορίας σε ποσοτική. Αυτή η μετατροπή ταυτίζεται με τη μετατροπή της πληροφορίας σε *δεδομένα* (data). Όταν η πληροφορία πάρει ποσοτικά χαρακτηριστικά μπορούμε να την επεξεργαστούμε στατιστικά²⁵.

Οι διαδικασίες μετατροπής της ποιοτικής πληροφορίας σε ποσοτικά δεδομένα είναι δύο.

Η πρώτη είναι η χρήση τακτικής κλίμακας, που μας επιτρέπει να κατατάξουμε μία απάντηση (για παράδειγμα κάτι που είπε ένας εξεταζόμενος με βάση κάποιο ερέθισμα – π.χ. φωτογραφία– και την οδηγία μίας δοκιμασίας) σε μία από τις ιεραρχημένες βαθμίδες μιας τακτικής κλίμακας. Πιο συγκεκριμένα ακόμη, ο εξεταζόμενος λέει κάτι και ο εξεταστής του προφορικού αποφασίζει αν αυτό που ειπώθηκε, ως προς την καταλληλότητα της

²⁵ Η διάκριση ποιοτικού και ποσοτικού είναι αντικείμενο που χρήζει παραπέρα εξέτασης αν πρόθεση κάποιου είναι η εμπάθηση λόγω προσωπικού ενδιαφέροντος. Δεδομένου του στόχου αυτού του άρθρου, αυτά που αναφέρθηκαν παραπάνω αρκούν. Κλείνοντας, ωστόσο, το θέμα πρέπει να διευκρινίσω ότι άλλο είναι η διχοτομία ποιοτική versus ποσοτική μεταβλητή, άλλο η διχοτομία ποιοτικά versus ποσοτικά δεδομένα και άλλο ακόμη η διχοτομία ποιοτική versus ποσοτική ανάλυση. Για την τελευταία αυτή διχοτομία, και ειδικότερα για την παρουσίαση τεχνικών ποιοτικής ανάλυσης, ο αναγνώστης παραπέμπεται στο δημοσίευμα που επιμελήθηκε ο H. Dorvil (2007: 409-445).

επιλογής του λεξιλογίου (ή ως προς κάποιαν άλλη μεταβλητή, όπως η συνοχή του λόγου ή η γραμματική ορθότητα, κ.τ.λ.), εντάσσεται στη βαθμίδα 0 ή 1 ή 2 ή 3. Τον αριθμό των βαθμίδων τον έχουν αποφασίσει άλλοι, αυτοί που οφείλουν επίσης να δώσουν όσο γίνεται πιο σαφείς οδηγίες στον εξεταστή/βαθμολογητή για το πότε να επιλέγει καθεμία από αυτές τις τέσσερις βαθμίδες. Αυτή η διαδικασία είναι ένα από τα τέσσερα είδη *μέτρησης* που έχει επινοήσει ο άνθρωπος. Αυτή η διαδικασία ενέχει πάντα κάποιο βαθμό υποκειμενικότητας.

Η δεύτερη διαδικασία μετατροπής ποιοτικής πληροφορίας σε ποσοτικά δεδομένα λέγεται *αρίθμηση* (counting) και συνίσταται στην καταγραφή του αριθμού φορών εμφάνισης κάποιου στοιχείου. Για παράδειγμα, σε μία δοκιμασία πολλαπλής επιλογής με δέκα *ερωτήματα* (items) ένας εξεταζόμενος απάντησε στα εννέα και άφησε ένα αναπάντητο. Ο ηλεκτρονικός υπολογιστής ή ένας βαθμολογητής σε πρώτη φάση θα ελέγξει, μηχανιστικά, αν η κάθε απάντηση ταυτίζεται με τη σωστή/αναμενόμενη και σε δεύτερη φάση θα αριθμήσει ή καταμετρήσει τις φορές που η απάντηση που δόθηκε ταυτίζεται με την αναμενόμενη²⁶. Αυτή η αρίθμηση θα δώσει μία ποσότητα πάνω στην οποία θα γίνει αμέσως ένας πολλαπλασιασμός με ένα συντελεστή για να προκύψει ο βαθμός που απέσπασε ο εξεταζόμενος. Αυτή η διαδικασία δεν ενέχει υποκειμενικότητα. Χωρίς αυτό να σημαίνει ότι δεν μπορούν να γίνουν λάθη.

Με όσο περισσότερη συνέπεια έχει γίνει η βαθμολόγηση, τόσο μεγαλύτερη βαθμολογική εγκυρότητα διαθέτει η όλη ερευνητική δραστηριότητα. Μια δεύτερη, ωστόσο, ιδιότητα της βαθμολόγησης που επηρεάζει την εγκυρότητα είναι η λογική, δηλαδή το βάσιμο της βαθμολόγησης, αφού η επιλογή του συντελεστή με τον οποίο πολλαπλασιάζουμε κάθε σωστή απάντηση²⁷ καθώς και ο αριθμός βαθμίδων που έχει η επιμέρους κλίμακα με την οποία βαθμολογούμε κάθε χαρακτηριστικό του λόγου του εξεταζόμενου, πρέπει να υπακούουν σε κάποια λογική και να εξασφαλίζουν ισορροπία στην τελική βαθμολογία.

Συμπερασματικά, τα βασικά ερωτήματα που επιτρέπουν να εκτιμήσουμε το βαθμό εγκυρότητας βαθμολόγησης που διαθέτει ένα σύστημα πιστοποίησης της γλωσσομάθειας είναι:

- 1) Έχει η επιλογή των δοκιμασιών γίνει με τέτοιο τρόπο ώστε δύο παρόμοια τεστ να δίνουν το ίδιο περίπου αποτέλεσμα αν δοθούν στους ίδιους εξεταζόμενους την ίδια περίπου χρονική στιγμή;
- 2) Αν το τεστ αποτελείται από κλειστά ερωτήματα (επιλογή ή συμπλήρωση στοιχείων, όπου η σωστή απάντηση είναι μία μόνο ή λίγες και συγκεκριμένες), η βαθμολογία στο μισό τεστ είναι περίπου ίδια με αυτήν στο άλλο μισό;

²⁶ Στα συστήματα πιστοποίησης που για να εξουδετερώσουν τον παράγοντα τύχη δίνουν και «αρνητική βαθμολογία», γίνεται φυσικά καταμέτρηση και των λανθασμένων απαντήσεων, καθώς και των απουσιών απάντησης.

²⁷ Αυτός ο πολλαπλασιασμός γίνεται πάντα, παρόλο ότι δεν το συνειδητοποιούμε: ακόμη κι αν ένα σύστημα πιστοποίησης δίνει μία μονάδα για κάθε σωστή απάντηση κλειστού τύπου, στην ουσία είναι σαν να έχει αποφασίσει ότι όλες οι σωστές απαντήσεις πολλαπλασιάζονται με το 1.

- 3) Αν το ίδιο τεστ δοθεί περισσότερες από μία φορές στους ίδιους εξεταζόμενους ή σε άλλους, που να διαθέτουν αποδεδειγμένα, κατά μέσον όρο, τον ίδιο βαθμό γλωσσομάθειας, δίνει περίπου το ίδιο αποτέλεσμα κάθε φορά που χρησιμοποιείται;
- 4) Αν το τεστ αποτελείται από ανοικτά ερωτήματα ή δοκιμασίες (ελεύθερη ή σχεδόν ελεύθερη παραγωγή λόγου, όπου η σωστή απάντηση δεν είναι μία αλλά πάρα πολλές, θεωρητικά μη πεπερασμένες σε αριθμό), η βαθμολόγηση του τεστ περισσότερες από μία φορές, από τον ίδιο βαθμολογητή σε διαφορετικές στιγμές, δίνει περίπου το ίδιο αποτέλεσμα;
- 5) Στην περίπτωση, και πάλι, των ανοικτών δοκιμασιών, η βαθμολόγηση της απάντησης των εξεταζομένων από διαφορετικούς βαθμολογητές, αδιάφορα με το αν είναι ταυτόχρονη ή ετερόχρονη, δίνει περίπου το ίδιο αποτέλεσμα; Κι αν το αποτέλεσμα είναι διαφορετικό, η διαφορά εμπίπτει σε ένα αποδεκτό εύρος διακύμανσης;
- 6) Πόσο βάσιμη και σωστά υπολογισμένη είναι η αναλογία μονάδων της τελικής βαθμολογίας που προκύπτουν από ανοικτά και κλειστά ερωτήματα ή που προέρχονται από καθεμία από τις *μακροδεξιότητες* (skills ή macro-skills) που ελέγχονται με το τεστ και τις ενότητες στις οποίες αυτό υποδιαιρείται;

2.5 Εγκυρότητα πρόβλεψης

Με τον όρο *εγκυρότητα πρόβλεψης* ή *πρόγνωσης* (predictive validity) γίνεται αναφορά στο βαθμό που τα αποτελέσματα της μέτρησης επιτρέπουν να προβλέψουμε αυτό που θα συμβεί στην πραγματική ζωή και συσχετίζονται θετικά με τα αποτελέσματα άλλων μετρήσεων, π.χ. στον εργασιακό ή κοινωνικό χώρο. Αυτές οι άλλες μετρήσεις ονομάζονται κριτήρια, γι' αυτό και η εγκυρότητα πρόβλεψης θεωρείται μία από τις διαστάσεις της εγκυρότητας που *ελέγχονται βάσει κριτηρίων* (criteria-related validity). Κατά τον έλεγχο της εγκυρότητας κριτήρια θεωρούνται οι μετρήσεις που αποτελούν σημείο αναφοράς και σύγκρισης.

Για να γίνει σαφέστερη η εγκυρότητα πρόβλεψης θα πάρουμε το παράδειγμα της πιστοποίησης της γλωσσομάθειας για ακαδημαϊκή χρήση. Όταν ένα πιστοποιητικό δηλώνει ότι ο κάτοχός του διαθέτει, για παράδειγμα, ικανοποιητική αγγλομάθεια ώστε να μπορεί να κάνει μεταπτυχιακές σπουδές σε κάποιο επιστημονικό κλάδο, περιμένουμε ο κάτοχός του να μπορεί να αντεπεξέλθει στις απαιτήσεις των σπουδών σε χρήση της αγγλικής γλώσσας. Έτσι μετράμε, πάνω σε ένα δείγμα κατόχων του πιστοποιητικού, την επιτυχία στις σπουδές, έτσι όπως αυτή απεικονίζεται ή αντιπροσωπεύεται στους βαθμούς που οι μεταπτυχιακοί φοιτητές απέσπασαν σε μεταπτυχιακά μαθήματα που απαιτούν τη χρήση της αγγλικής, π.χ. για βιβλιογραφική ενημέρωση. Η εγκυρότητα πρόβλεψης που διαθέτει το σύστημα πιστοποίησης που χορήγησε τα πιστοποιητικά προσδιορίζεται υπολογίζοντας το *δείκτη* ή *συντελεστή συσχέτισης* ή *συνάφειας* (coefficient of correlation) που προκύπτει από τη σύγκριση των αποτελεσμάτων του τεστ που χρησιμοποιεί το σύστημα και του μέσου όρου των αποτελεσμάτων σε συγκεκριμένα μαθήματα.

Είναι αυτονόητο ότι στην πιστοποίηση της γλωσσομάθειας γενικώς, δηλαδή για γενική χρήση, ο εντοπισμός και ο υπολογισμός κριτηρίου ή κριτηρίων, και της συνάφειάς τους με τη βαθμολογία στο τεστ, είναι ιδιαίτερα προβληματική. Όσο πιο περιορισμένη είναι η δομή, τόσο ευκολότερο είναι να ελέγξει κανείς την εγκυρότητα πρόβλεψης.

Τα σημαντικότερα ερωτήματα που τίθενται κατά τον έλεγχο της εγκυρότητας πρόβλεψης σχετίζονται με το κριτήριο. Ακολουθούν μερικά ερωτήματα:

- 1) Πόσο λογικά έγινε η επιλογή του κριτηρίου; Δηλαδή, υπάρχει πραγματικά σχέση της δομής και της επίδοσης ή δραστηριότητας, κ.τ.λ. την οποία το σύστημα πιστοποίησης χαρακτηρίζει κριτήριο για να ελέγξει την εγκυρότητα πρόβλεψης;
- 2) Μεσολάβησε αρκετός χρόνος μεταξύ της πρώτης μέτρησης (αξιολόγηση για την απόκτηση του πιστοποιητικού) και της δεύτερης μέτρησης (μέτρηση της ιδιότητας που θεωρείται κριτήριο), έτσι ώστε η μία (ετοιμασία για αντιμετώπιση του γλωσσικού τεστ και συμπλήρωσή του) να μην επηρεάσουν τη δεύτερη;
- 3) Πόσο αντιπροσωπευτική του πληθυσμού (δηλαδή του συνόλου των ατόμων που εξετάστηκαν από το σύστημα πιστοποίησης) είναι η ομάδα ατόμων επί της οποίας γίνεται η δεύτερη μέτρηση για εξασφάλιση του κριτηρίου;²⁸
- 4) Έγινε ο έλεγχος της συσχέτισης της επίδοσης στο γλωσσικό τεστ και του κριτηρίου με τρόπο στατιστικά θεμιτό; Δηλαδή υπολογίστηκε ο κατάλληλος δείκτης, δεδομένης της κλίμακας μέτρησης που χρησιμοποιήθηκε για την πιστοποίηση και για τη μέτρηση του κριτηρίου;²⁹
- 5) Μήπως το μη ικανοποιητικό αποτέλεσμα της συσχέτισης, μετά τον έλεγχο των παραπάνω ερωτημάτων, δεν οφείλεται στην έλλειψη εγκυρότητας πρόβλεψης, αλλά στην ανεπάρκεια ή την ακαταλληλότητα του κριτηρίου; Δηλαδή μήπως πρέπει να ξαναγίνει ο έλεγχος με άλλο κριτήριο ή, καλύτερα, με ένα συνδυασμό κριτηρίων;

2.6 Εγκυρότητα όψης

Η *εγκυρότητα όψης* ή *επιφάνειας* (face validity) είναι ο βαθμός στον οποίο οι «χρήστες» του συστήματος πιστοποίησης (βασικά οι εξεταζόμενοι ή υποψήφιοι, οι εκπαιδευτικοί που τους ετοιμάζουν και οι εργοδότες, οι ακαδημαϊκοί δάσκαλοι, κ.τ.λ. στους οποίους οι κάτοχοι των πιστοποιητικών τα καταθέτουν) θεωρούν, «εκ πρώτης όψεως», ότι το σύστημα είναι έγκυρο, δηλαδή μετράει αξιόπιστα, σταθερά, δίκαια τη γλωσσομάθεια και όχι κάτι άλλο. Στην πραγματικότητα η εγκυρότητα όψης είναι ένας συνδυασμός όλων των προηγούμενων διαστάσεων έτσι όπως τις κρίνουν, κατά προσέγγιση, μη ειδικοί στον έλεγχο της εγκυρότητας.

²⁸ Το ερώτημα αυτό τίθεται σχεδόν πάντα, αφού είναι πρακτικά απίθανο να μπορούμε να κάνουμε τη δεύτερη μέτρηση πάνω σε όλα τα υποκείμενα. Συνεπώς η δειγματοληψία έχει συνέπειες στο αποτέλεσμα του ελέγχου της εγκυρότητας πρόβλεψης.

²⁹ Το είδος της κλίμακας που χρησιμοποιούμε προσδιορίζεται από τη φύση της μεταβλητής και προσδιορίζει, στη συνέχεια, το συντελεστή/δείκτη συσχέτισης/συνάφειας που μπορούμε και πρέπει να υπολογίσουμε.

Αυτή η διάσταση της εγκυρότητας έχει δεχτεί κριτική από πολλούς ειδικούς ή και έχει γίνει αντικείμενο απόρριψης, ως μη επιστημονικό κριτήριο αξιολόγησης ενός συστήματος πιστοποίησης.

Συμφωνώντας με την Anastasi (1988: 144), θα θεωρήσω ότι η εγκυρότητα όψης είναι κεφαλαιώδους σημασίας για τα συστήματα πιστοποίησης, αφού χωρίς αυτήν οι εξεταζόμενοι δε θα αφιερώσουν το χρόνο και δε θα κάνουν την προσπάθεια που απαιτούνται για τη διεξαγωγή των δοκιμασιών του τεστ, με αποτέλεσμα να αλλοιώνονται τα αποτελέσματα, παρά τον ενδεχόμενα υψηλό βαθμό εγκυρότητας δομής, περιεχομένου, περιστασης, βαθμολόγησης και πρόβλεψης του συστήματος.

Επίσης, ένα σύστημα πιστοποίησης δε διεξάγει ακαδημαϊκή έρευνα, δηλαδή έρευνα που οφείλει να υπακούει μόνο στις νόρμες επιστημονικού ήθους και της μεθοδολογίας της έρευνας, αλλά εφαρμοσμένη έρευνα προσανατολισμένη στη λήψη αποφάσεων, που οφείλει να λαμβάνει υπόψη εξωτερικά τιθέμενες δεσμεύσεις (constrains). Ο H. Braun χαρακτηρίζει, σωστά, την ετοιμασία ενός τεστ και το σχεδιασμό μιας έρευνας του είδους που μας απασχολεί εδώ ως επιδίωξη του «καλύτερου δυνατού σχεδίου, δεδομένων των δεσμεύσεων»³⁰ (Braun, 2000: 263), κάνοντας τον παραλληλισμό αυτής της εργασίας με τον αρχιτεκτονικό σχεδιασμό, κατά τον οποίο ο αρχιτέκτονας είναι υποχρεωμένος να λάβει υπόψη τις επιθυμίες αυτού που του ανέθεσε τη δουλειά, τις απόψεις του πολιτικού μηχανικού για το πόσο υλοποιήσιμο είναι ένα σχέδιο, δεδομένων των υλικών που υπάρχουν στην αγορά και της τεχνογνωσίας που διαθέτουν αυτοί που θα κάνουν την κατασκευή, κ.τ.λ. Αν αρνηθεί να προσαρμόσει τη σχεδιαστική του έμπνευση στις εξωτερικές δεσμεύσεις, διατρέχει τον κίνδυνο να μείνει το σχέδιό του στα χαρτιά. Αντίστοιχα, αν ένα σύστημα πιστοποίησης αδιαφορήσει για την εγκυρότητα όψης, διατρέχει τον κίνδυνο να μην έχει ενδιαφερόμενους να εξεταστούν, δηλαδή να οδηγηθεί σε αυτοκατάργηση.

Τα ερωτήματα που δίνουν το μέτρο εγκυρότητας όψης ενός συστήματος πιστοποίησης της γλωσσομάθειας είναι πολλά, μερικά εκ των οποίων καταγράφονται παρακάτω:

- 1) Τι σκέφτονται όσοι εξετάστηκαν και οι υποψήφιοι για την απόκτηση του πιστοποιητικού αναφορικά με το επίπεδο δυσκολίας της εξέτασης σε σχέση με τα άλλα συστήματα που είναι στη διάθεσή τους;
- 2) Τι σκέφτονται αναφορικά με τη σοβαρότητα της εξέτασης, δηλαδή με το πόσο σχετικές είναι οι εξετάσεις με αυτό που πιστοποιείται;
- 3) Τι σκέφτονται για το δίκαιο και το ακριβές της βαθμολογίας;
- 4) Κατά πόσο θεωρούν ότι το πιστοποιητικό έχει αναγνώριση στη χώρα τους και στο εξωτερικό για συνέχιση σπουδών ή για εξασφάλιση θέσης εργασίας, κ.τ.λ.;
- 5) Όσοι εξετάστηκαν, θα επεδίωκαν και πάλι να πάρουν μέρος στις εξετάσεις αν απορρίφθηκαν ή αν χρειάζονταν πιστοποιητικό υψηλότερου επιπέδου από αυτό που απέκτησαν, με βάση τα παραπάνω;

³⁰ One can think of building a test as a problem that falls under the rubric of 'optimal design under constraints'.

2.7 Εγκυρότητα συνεπειών

Τα ίδια τα αποτελέσματα μιας αξιολογικής διαδικασίας, οι μέθοδοι που εφαρμόστηκαν και η ερμηνεία των αποτελεσμάτων από τρίτους συχνά έχουν συνέπειες για τους ίδιους τους εξεταζόμενους ή άλλα άτομα, για το σχολείο ή άλλους θεσμούς και ιδρύματα και για την κοινωνία στο σύνολό της. Οι συνέπειες αυτές, στην περίπτωση των συστημάτων πιστοποίησης της γλωσσομάθειας επιβάλλεται να είναι θετικές, ή τουλάχιστον ουδέτερες. Οι αρνητικές συνέπειες, στις οποίες συνήθως γίνεται αναφορά με τον όρο washback ή backwash effect, δηλαδή η αρνητική επίδραση ή επίπτωση (impact) της εξέτασης στην καθημερινή σχολική πρακτική (βλ. και το άρθρο της Τσαγγαρή σε αυτόν εδώ τον ιστότοπο), μπορεί να αφαιρέσουν βαθμούς από την εγκυρότητα ως ενιαία δομή/ιδιότητα των ερευνών αυτού του είδους.

Το πόσο θετικές/επιθυμητές ή αρνητικές/ανεπιθύμητες είναι οι συνέπειες προσδιορίζει το βαθμό εγκυρότητας συνεπειών (consequential validity).

Την προσοχή της διεθνούς επιστημονικής κοινότητας σε αυτή τη διάσταση της εγκυρότητας τράβηξαν, όπως ήδη αναφέρθηκε, οι Messick και Bachman. Ορισμένοι αρνούνται να αποδεχτούν πως οι συνέπειες, ιδιαίτερα όταν οφείλονται σε διαφοροποιημένη ερμηνεία που τρίτοι κάνουν, είναι επιτρεπτό να προσδιορίσουν το βαθμό εγκυρότητας του συστήματος πιστοποίησης.

Υιοθετώντας για λίγο τη θέση τους, θα δώσω ένα παράδειγμα. Το Κρατικό Πιστοποιητικό Γλωσσομάθειας (ΚΠΓ) ελέγχει την ικανότητα *διαμεσολάβησης* (mediation). Διαμεσολάβηση, σύμφωνα με το ΚΕΠΑ (2001: 14), έχουμε όταν «τόσο ως προς την αντίληψη όσο και ως προς την παραγωγή, οι γραπτές ή/και προφορικές δραστηριότητες **διαμεσολάβησης** καθιστούν δυνατή την επικοινωνία μεταξύ ανθρώπων που για κάποιο λόγο δεν είναι δυνατόν να επικοινωνήσουν απευθείας μεταξύ τους. Η μετάφραση ή η διερμηνεία, μια παράφραση, μια περίληψη ή αναφορά, παρέχουν σε τρίτους μια (ανα)διατύπωση ενός πρωτότυπου κειμένου στο οποίο δεν έχουν οι ίδιοι άμεση πρόσβαση.»³¹. Από αυτά τα είδη διαμεσολάβησης –που άλλωστε είναι μόνο παραδείγματα και υπάρχουν κι άλλα– το ΚΠΓ αποφάσισε, στις προδιαγραφές του, να αποκλείσει τη μετάφραση και τη διερμηνεία. Παρά το σαφή αποκλεισμό αυτών των μορφών διαμεσολάβησης, υπάρχουν πληροφορίες ότι σε αρκετά Κέντρα Ξένων Γλωσσών οι καθηγητές ασκούν τους υποψήφιους στη μετάφραση. Αυτό μπορεί να χαρακτηριστεί αρνητική επίδραση του ΚΠΓ πάνω στην καθημερινή σχολική πράξη αυτών των Κέντρων.

³¹ Η απόδοση στην ελληνική είναι από τη μετάφραση του ΚΕΠΑ που έκαναν συνεργάτες του Κέντρου Ελληνικής Γλώσσας. Στην αγγλική το απόσπασμα είναι: In both the receptive and productive modes, the written and/or oral activities of **mediation** make communication possible between persons who are unable, for whatever reason, to communicate with each other directly. Translation or interpretation, a paraphrase, summary or record, provides for a third party a (re)formulation of a source text to which this third party does not have direct access.

Κατά την άποψη των ειδικών που δε «χρεώνουν» το εκάστοτε σύστημα πιστοποίησης για τις λανθασμένες ερμηνείες που κάνουν τρίτοι, το ΚΠΓ δεν έχει καμία ευθύνη, και αυτό δεν πρέπει να θεωρηθεί πως μειώνει την εγκυρότητα του συστήματος.

Ένα δεύτερο απτό παράδειγμα. Το επίπεδο Γ2 ορισμένων συστημάτων πιστοποίησης αναγνωρίζεται από το Υπουργείο Παιδείας ΔΒΜΘ ως τίτλος που εξασφαλίζει το δικαίωμα διδασκαλίας της ξένης γλώσσας. Έτσι, με την επίδειξη του πιστοποιητικού γλωσσομάθειας ή γλωσσικής επάρκειας, το Υπουργείο Παιδείας χορηγεί ένα πιστοποιητικό «διδασκτικής επάρκειας». Αυτό αποτελεί λανθασμένη «ερμηνεία» του αποτελέσματος της όλης διαδικασίας πιστοποίησης της γλωσσομάθειας.

Κατά την άποψη των ίδιων ειδικών, τα συστήματα που είναι στην κατάσταση του Υπουργείου Παιδείας για τη χορήγηση της επάρκειας διδασκαλίας δεν έχουν ευθύνη για τις αποφάσεις του Υπουργείου.

Αλλάζοντας πλευρά ή θέση –και υιοθετώντας την άποψη του Bachman (1990: 279) ότι «...τα τεστ δεν ετοιμάζονται και δε χρησιμοποιούνται σε έναν ψυχομετρικό δοκιμαστικό σωλήνα χωρίς αξίες, αλλά ιδεατά σχεδόν πάντα προορίζονται για να ικανοποιήσουν τις ανάγκες ενός εκπαιδευτικού συστήματος ή μιας κοινωνίας γενικότερα»³²– θα ισχυριστώ ότι το ΚΠΓ, στην πρώτη περίπτωση, έχει ευθύνη και πρέπει να δράσει με κάποιον ή κάποιους από τους παρακάτω τρόπους: α) να κάνει σεμινάρια στους καθηγητές, β) να κάνει ακόμη σαφέστερες τις προδιαγραφές, γ) να αλλάξει τις δοκιμασίες έτσι ώστε να είναι αδύνατο ή ανεπίτρεπτο να μεταφράσει ο εξεταζόμενος, ή, αν τα παραπάνω μέτρα δεν αποδώσουν και το πρόβλημα έχει μεγάλη έκταση, δ) να καταργήσει τις δοκιμασίες γραπτής και προφορικής διαμεσολάβησης. Η τελευταία επιλογή θα αύξανε την εγκυρότητα συνεπειών και θα μείωνε την εγκυρότητα δομής του συστήματος. Το θέμα αυτό, δηλαδή το θέμα της αλληλεξάρτησης των διαστάσεων της εγκυρότητας, εξετάζεται και στην αρχή του επόμενου κεφαλαίου.

Επίσης, τα συστήματα των οποίων τα πιστοποιητικά εξασφαλίζουν επάρκεια διδασκαλίας έχουν ευθύνη και πρέπει να κάνουν επίσημα γνωστό στο ΥΠΕΠΘ ότι γίνεται στρεβλή ερμηνεία του αποτελέσματος της μέτρησης που αυτά έκαναν.

Ανακεφαλαιώνοντας, μπορεί να πει κανείς ότι μερικά βασικά ερωτήματα που πρέπει να βρουν απάντηση κατά τον έλεγχο της εγκυρότητας συνεπειών είναι:

- 1) Μήπως το περιεχόμενο του τεστ ή οι τύποι δοκιμασιών που χρησιμοποιεί ή κάποιο άλλο χαρακτηριστικό του επηρεάζουν –αρνητικά ή θετικά– τη συμπεριφορά των διδασκόντων, επιβάλλοντάς τους, έμμεσα, στρεβλές ή επιθυμητές σχολικές πρακτικές;
- 2) Μήπως όσοι ετοιμάζονται για συμμετοχή στις εξετάσεις του συστήματος πιστοποίησης υποχρεώνονται έμμεσα από αυτό σε ανάπτυξη στρεβλών (για

³² ...tests are not developed and used in value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of a society at large.

παράδειγμα επινόηση «κόλπων» για εξασφάλιση επιτυχίας στην εξέταση) ή επιθυμητών μαθησιακών στρατηγικών;

- 3) Μήπως υπάρχουν περιπτώσεις ομάδων εξεταζομένων ή μεμονωμένων ατόμων που για κάποιο λόγο (πολιτισμικό υπόβαθρο, ενδυμασία, ηλικία, χρώμα του δέρματος, κ.τ.λ.) βρέθηκαν σε μειονεκτική θέση και αδικήθηκαν, με αποτέλεσμα να έχουν συνέπειες (ψυχολογικές, στην εργασία τους, στις σπουδές τους, κ.τ.λ.); Και αν ναι, πόσο συστηματικό ή πόσο περιστασιακό είναι αυτό το φαινόμενο;
- 4) Τα χαρακτηριστικά του τεστ και του συστήματος πιστοποίησης γενικότερα έχουν θετικές (ή μήπως αρνητικές) συνέπειες πάνω στη βιομηχανία παραγωγής διδακτικών και μαθησιακών υλικών, σε συμβατική ή/και σε ηλεκτρονική μορφή;
- 5) Η λειτουργία του συστήματος πιστοποίησης έχει θετική, αρνητική ή δεν έχει καμία συνέπεια στον τρόπο λήψης αποφάσεων εκ μέρους των αρμόδιων γι' αυτό οργανισμών ή φορέων (π.χ., στην περίπτωση της Ελλάδας, Υπουργείο Παιδείας, συνδικαλιστικοί φορείς, κ.τ.λ.);
- 6) Τι είδους συνέπειες προκαλεί το σύστημα πιστοποίησης πάνω στα άλλα συστήματα πιστοποίησης που είναι στη διάθεση των ίδιων εξεταζομένων, δηλαδή που «μοιράζονται την ίδια αγορά»;

3. Έλεγχος της εγκυρότητας

Όπως πρέπει να έχει ήδη γίνει κατανοητό, δεδομένου ότι η εγκυρότητα έχει διάφορες πτυχές ή διαστάσεις, ο έλεγχος της εγκυρότητας δεν μπορεί παρά να καταλήγει σε αποτύπωση του *βαθμού* (degree) της εγκυρότητας. Δηλαδή η εγκυρότητα δεν είναι άσπρο-μαύρο. Μια έρευνα μπορεί να διαθέτει υψηλό βαθμό εγκυρότητας όψης και χαμηλό βαθμό εγκυρότητας δομής. Επίσης, μπορεί να συμβεί, στην προσπάθεια αύξησης της εγκυρότητας βαθμολόγησης, ένα σύστημα πιστοποίησης γλωσσομάθειας να χάσει σε εγκυρότητα δομής και αντίστροφα.

Για να γίνει αυτό σαφέστερο θα πάρουμε το παράδειγμα ενός συστήματος που διαπιστώνει ότι στις ενότητες της εξέτασης όπου οι εξεταζόμενοι μπορούν να απαντήσουν σχετικά ελεύθερα (ερωτήματα «ανοικτά» ή «αποκλίνουσας παραγωγής λόγου») οι βαθμολογητές επηρεάζονται από τη νόρμα και τις προσωπικές προσδοκίες τους, με συνέπεια τη διατήρηση χαμηλού βαθμού εγκυρότητας βαθμολόγησης (δηλαδή, βασικά, ενδο- και δια-βαθμολογικής αξιοπιστίας). Έτσι μειώνει ή καταργεί αυτού του είδους τα ερωτήματα. Αυτό θα οδηγήσει αναπόφευκτα σε αύξηση του βαθμού εγκυρότητας βαθμολόγησης και σε μείωση της εγκυρότητας δομής, αφού η γλωσσομάθεια δεν προϋποθέτει ή δεν έχει ως εξωτερική εκδήλωση (μόνο) την ικανότητα επιλογής μεταξύ λίγων γλωσσικών στοιχείων (ερωτήματα πολλαπλής επιλογής, ναι/όχι, κ.τ.λ.).

Το πολυδιάστατο της εγκυρότητας, σε συνδυασμό με το γεγονός ότι κάθε μία από τις διαστάσεις της δεν έχει σε κάθε περίπτωση γλωσσικής αξιολόγησης την ίδια σημασία ή το

ίδιο ειδικό βάρος, κάνει τον έλεγχο της ολικής/σφαιρικής εγκυρότητας αφενός θέμα βαθμού και αφετέρου θέμα προσωπικής εκτίμησης. Από όσο είμαι σε θέση να γνωρίζω δεν έχει ακόμη επινοηθεί και γίνει κοινά αποδεκτό ένα σύστημα ελέγχου της εγκυρότητας που να καταλήγει σε σαφές και ποσοτικά εκφρασμένο αποτέλεσμα. Το γεγονός αυτό κάνει πολύ δύσκολο το εγχείρημα έγκυρης αξιολόγησης των συστημάτων πιστοποίησης της γλωσσομάθειας με προοπτική τη *διαπίστευσή* τους (accreditation).

Η θέση που διατυπώνεται στην προηγούμενη παράγραφο, ότι δηλαδή κάθε μία από τις διαστάσεις της εγκυρότητας δεν έχει το ίδιο βάρος με τις άλλες, δε φαίνεται να βρίσκει σύμφωνους όλους τους ειδικούς. Αντίθετα οι περισσότεροι ισχυρίζονται ότι όλοι οι τύποι εγκυρότητας έχουν την ίδια αξία. Για παράδειγμα ο Weir (2005: 13) γράφει: «Κανένα είδος εγκυρότητας δεν πρέπει να θεωρείται ανώτερο από ένα άλλο»³³. Θεωρώ πως μια τέτοια άποψη έρχεται σε αντίθεση με την γενικά αποδεκτή ιδέα ότι για τον έλεγχο της εγκυρότητας πρέπει να λαμβάνεται υπόψη ο λόγος για τον οποίο γίνεται η γλωσσική αξιολόγηση ή, διαφορετικά, ποια είναι η *λειτουργία* (function) της αξιολόγησης, καθώς και το περιβάλλον στο οποίο εντάσσεται. Μερικές από τις λειτουργίες της αξιολόγησης και κάποιες τοπικές ιδιαιτερότητες μπορεί να προσδίδουν μεγαλύτερο βάρος σε μία διάσταση της εγκυρότητας από ό,τι σε κάποιαν άλλη. Για παράδειγμα, για ένα σύστημα πιστοποίησης της γλωσσομάθειας του οποίου τα πιστοποιητικά αναγνωρίζονται από το ελληνικό κράτος και επιτρέπουν απόκτηση «επάρκειας διδασκαλίας», η εγκυρότητα βαθμολόγησης, για παράδειγμα, και η εγκυρότητα συνεπειών αποκτούν μεγαλύτερη σημασία στην Ελλάδα από την αξία που οι ίδιες διαστάσεις έχουν σε άλλη χώρα, αν εκεί τα πιστοποιητικά του ίδιου συστήματος δεν αναγνωρίζονται. Και αυτό δε σημαίνει πως ταυτόχρονα αποκτά, στον ίδιο βαθμό, μεγαλύτερη αξία και η εγκυρότητα περιεχομένου. Συνεπώς, η κάθε διάσταση έχει λίγο διαφορετικό βάρος, ανάλογα με την περίπτωση ή και ανάλογα με το επίπεδο γλωσσομάθειας που πιστοποιείται, στην ίδια περίπτωση. Κι αυτό είναι θέμα εκτίμησης αυτού που κάνει τον έλεγχο.

3.1 Η φύση του ελέγχου εγκυρότητας

Στην αρχή αυτού του άρθρου διατυπώθηκε η άποψη πως η αξιολόγηση της γλωσσομάθειας με στόχο την πιστοποίηση είναι μία μορφή επιστημονικής έρευνας. Είναι αυτονόητο ότι και ο έλεγχος της εγκυρότητας των συστημάτων πιστοποίησης είναι ερευνητική δραστηριότητα. Πρόκειται για έρευνα που απαιτεί διάφορες μετρήσεις, οι οποίες επιτρέπουν να ελεγχθεί μία σειρά από υποθέσεις, από τις οποίες οι τρεις κεντρικότερες είναι πως: α) «η γλωσσική συμπεριφορά του εξεταζόμενου –δηλαδή οι απαντήσεις του στις δοκιμασίες στις οποίες υποβλήθηκε– αποτελούν εξωτερική εκδήλωση ή (εν)δείκτη (indicator) της δομής που μετράται»³⁴, β) «η γλωσσική αυτή συμπεριφορά ποσοτικοποιείται, δηλαδή μεταφράζεται

³³ No single validity can be considered superior to another.

³⁴ Η επαλήθευση αυτής της πρώτης ερευνητικής υπόθεσης δίνει υψηλό βαθμό εγκυρότητας δομής, περιεχομένου, περίπτωσης και πρόβλεψης.

σε βαθμούς, ή βαθμολογία, με τρόπο αξιόπιστο και δίκαιο»³⁵ και γ) «το αποτέλεσμα της όλης διαδικασίας ερμηνεύεται σωστά και έχει τις πρέπουσες συνέπειες πάνω στα άτομα και στην κοινωνία»³⁶.

Ο έλεγχος της εγκυρότητας είναι η διαρκής προσπάθεια ελέγχου των ερευνητικών υποθέσεων και της εξασφάλισης των πειστηρίων που αποδεικνύουν το βαθμό αλήθειας αυτών των υποθέσεων³⁷. Ο έλεγχος ουσιαστικά –και συχνά υποσυνείδητα– αρχίζει με τον ίδιο το σχεδιασμό του συστήματος, πριν αυτό λειτουργήσει.

3.2 Οι φορείς που οφείλουν να ελέγχουν την εγκυρότητα

Ο έλεγχος της εγκυρότητας πρέπει να γίνεται από τα ίδια τα συστήματα πιστοποίησης. Είναι ένας αυτοέλεγχος που τους επιτρέπει να βελτιώνονται ή να διατηρούν έναν υψηλό βαθμό εγκυρότητας, που, όπως ήδη έχει αναφερθεί, είναι η βασική συνισταμένη της ποιότητας της ερευνητικής δραστηριότητας, και επομένως του κάθε συστήματος.

Τα σοβαρά συστήματα έχουν προβλέψει και υλοποιούν τέτοιο έλεγχο. Τέτοια συστήματα, όμως, υπάρχουν λίγα. Τα περισσότερα συστήματα –αν κρίνει κανείς από στοιχεία που τα ίδια δημοσιοποιούν– κάνουν μερικό μόνο ή και καθόλου έλεγχο.

Αν ο ανεπαρκής έλεγχος οφείλεται σε αδυναμία του συστήματος να αυτοελεγχθεί, τότε το σύστημα πρέπει αυτονόητα να απευθυνθεί σε εξωτερικό αξιολογητή, που μπορεί να είναι άτομο/-α ή, καλύτερα, φορέας.

Επειδή η κοινωνία και η πολιτεία, ιδιαίτερα στην Ελλάδα, αποδίδουν σημαντική αξία στην πιστοποίηση της γλωσσομάθειας, δε θεωρώ επιτρεπτό ο έλεγχος της εγκυρότητας να επαφίεται στη βούληση των ίδιων των συστημάτων πιστοποίησης. Συνεπώς κρίνεται απαραίτητο ο έλεγχος αυτός να γίνεται και από επίσημο κρατικό φορέα. Ακόμη καλύτερα θα ήταν αν τον έλεγχο αναλάμβανε διεθνής επίσημος φορέας, τουλάχιστον σε επίπεδο Ευρωπαϊκής Ένωσης, αφού τα περισσότερα συστήματα πιστοποίησης από αυτά που είναι στη διάθεση των ελλήνων πολιτών έχουν διεθνή χαρακτήρα.

3.3 Οι μέθοδοι ελέγχου της εγκυρότητας

Κατά τον έλεγχο της εγκυρότητας υπάρχουν δύο βασικοί τρόποι εργασίας. Ο ένας συνίσταται στην επεξεργασία ποιοτικών και ποσοτικών δεδομένων που προκύπτουν από την ίδια τη λειτουργία του συστήματος: προδιαγραφές, γλωσσική παραγωγή των εξεταζομένων, βαθμολογία σε κάθε δοκιμασία, κ.τ.λ. Ο έλεγχος που στηρίζεται σε τέτοιου

³⁵ Η επαλήθευση της δεύτερης βασικής ερευνητικής υπόθεσης δίνει υψηλό βαθμό εγκυρότητας βαθμολόγησης.

³⁶ Η επαλήθευση αυτής της τελευταίας ερευνητικής υπόθεσης δίνει, προφανώς, υψηλό βαθμό εγκυρότητας όψης και συνεπειών.

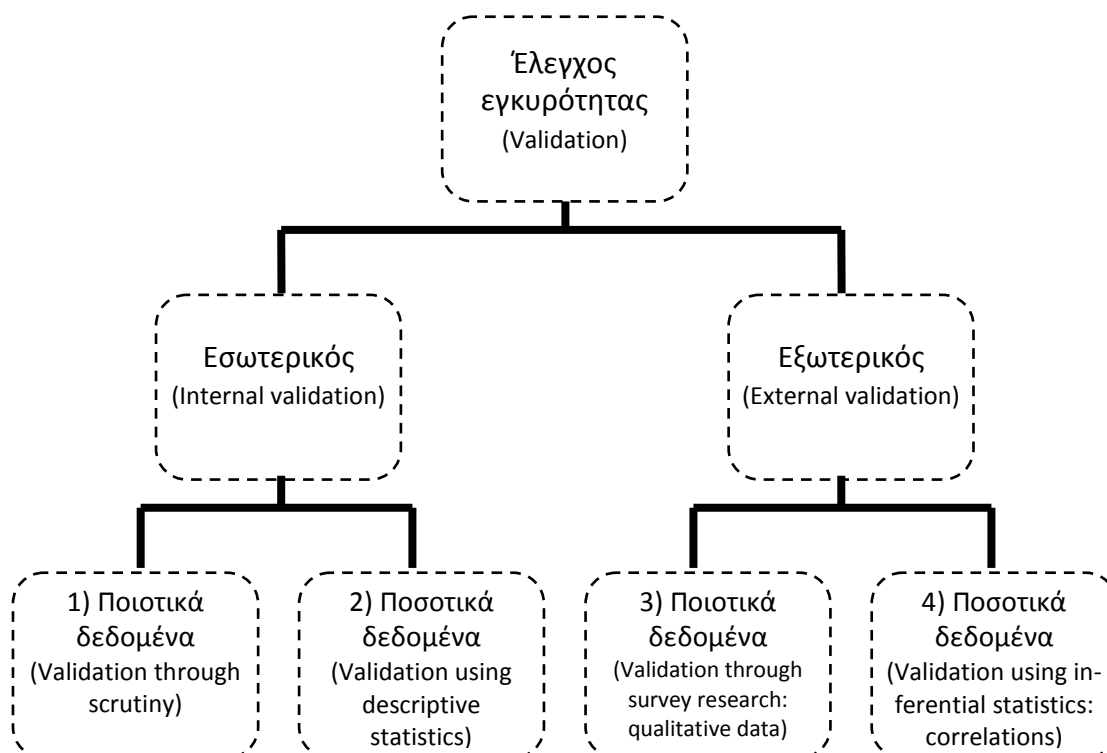
³⁷ Όπως είναι γνωστό οι ερευνητικές υποθέσεις άλλοτε διατυπώνονται με μορφή καταφατικής πρότασης και άλλοτε με μορφή ερώτησης, ανάλογα με το πόση έρευνα έχει προϋπάρξει πάνω σε ένα πεδίο ή αντικείμενο μελέτης. Με άλλα λόγια οι τρεις βασικές ερευνητικές υποθέσεις που διατυπώνονται εδώ με μορφή καταφατικών προτάσεων δεν είναι κάτι τελείως διαφορετικό από τις ερωτήσεις που καταγράφονται παραπάνω στο τέλος της εξέτασης κάθε διάστασης της εγκυρότητας (πρβ. Berthier, 2006³: 43-46).

είδους στοιχεία μπορεί να ονομαστεί *εσωτερικός έλεγχος* (internal validation). *Εξωτερικός έλεγχος* (external validation) θεωρείται εδώ ο έλεγχος που απαιτεί νέα δεδομένα, ποιοτικά ή/και ποσοτικά που πρέπει να εξασφαλιστούν από αυτόν που κάνει τον έλεγχο αναζητώντας τα έξω από το σύστημα πιστοποίησης.

Ο εσωτερικός έλεγχος επιτυγχάνεται μέσω δύο κυρίως μεθόδων: α) την εφαρμογή λογικού συμπερασμού, που μόνο ειδήμονες διαφόρων κλάδων σχετικών με τη γλωσσομάθεια μπορούν να κάνουν (θεωρητικός ή ποιοτικός έλεγχος) και β) την εφαρμογή στατιστικών μεθόδων, δηλαδή ανάλυσης ποσοτικών δεδομένων που προέκυψαν από μετρήσεις (στατιστικός ή ποσοτικός έλεγχος).

Στην προσπάθεια εξασφάλισης συμμετρίας που διευκολύνει την απομνημόνευση, και απλοποιώντας κάπως τα πράγματα, θα θεωρήσουμε ότι και ο εξωτερικός έλεγχος επιτυγχάνεται, επίσης, μέσω δύο μεθόδων: α) τη διεξαγωγή δειγματοληπτικών ερευνών που εξασφαλίζουν ποιοτικά δεδομένα από έξω από το σύστημα και β) την εξασφάλιση ποσοτικών δεδομένων που επιτρέπουν συσχέτιση με τα ποσοτικά δεδομένα που παρέχει το ίδιο το σύστημα πιστοποίησης.

Για τον έλεγχο καθεμίας από τις διαστάσεις της εγκυρότητας είναι πιθανό να χρειαστεί να εφαρμόσει κανείς περισσότερες από μία μεθόδους. Κάθε διάσταση, ωστόσο, από τη φύση της, αξιοποιεί περισσότερο κάποια από τις μεθόδους από ό,τι τις άλλες. Το σχήμα που ακολουθεί κάνει εποπτική και ευκολότερα κατανοητή τη θέση που διατυπώνεται στις πέντε προηγούμενες παραγράφους.



Σχήμα 2: Απεικόνιση των διαδικασιών ελέγχου της εγκυρότητας

Η εξέταση καθεμίας από τις παραπάνω διαδικασίες με λεπτομέρεια και δίνοντας συγκεκριμένα παραδείγματα θα απαιτούσε πολλές σελίδες κειμένου. Υπάρχουν συγγράμματα αφιερωμένα ολόκληρα σε κάθε μία από τις διαδικασίες³⁸. Θα περιοριστούμε, συνεπώς, σε μία συνοπτική και εν πολλοίς επιφανειακή εξέταση των διαδικασιών.

Η πρώτη από τις παραπάνω διαδικασίες βρίσκει εφαρμογή κατά τον έλεγχο, κυρίως, της εγκυρότητας δομής και περιεχομένου. Αυτό ισχύει φυσικά για τη διάσταση της εγκυρότητας έτσι όπως την ορίσαμε παραπάνω, αφού η εγκυρότητα δομής, κατά την άποψη των περισσότερων σύγχρονων δοκιμασιολόγων, έχει πολύ ευρύτερο αναφερόμενο και απαιτεί έλεγχο και μέσω της τέταρτης διαδικασίας.

Ως παράδειγμα εφαρμογής της πρώτης διαδικασίας θα πάρουμε την περίπτωση που κάποιος, για να απαντήσει στο ερώτημα «ποιο θεωρητικό πλαίσιο υιοθετείται από το σύστημα και πόσο ομοιογενές είναι αυτό», χρειάζεται να μελετήσει τεκμήρια που το σύστημα δίνει και να εκτιμήσει αν τα δηλούμενα έχουν σαφήνεια και ομοιογένεια, δεδομένης της επιστημονικής γνώσης και των επικρατούσων θεωριών, τη στιγμή του ελέγχου, στις επιστήμες που μελετούν τις φυσικές γλώσσες, τη γλωσσική επικοινωνία και την *εκμάθηση* (learning) ή *πρόσκτηση* (acquisition) ενός γλωσσικού κώδικα.

Το βάσιμο του τρόπου μετατροπής της δομής σε έναν πεπερασμένο αριθμό μεταβλητών που (θα) ελέγχονται με τις δοκιμασίες, απαιτεί επίσης ανάλυση ποιοτικών δεδομένων σε συνεργασία και με ειδικούς στη μετρολογία και στη γλωσσική δοκιμασιολογία. Και σε αυτή την περίπτωση το πόρισμα αναφορικά με το βαθμό εγκυρότητας εξαρτάται από τις επικρατούσες θεωρίες στις επιστήμες αυτές.

Η δεύτερη από τις διαδικασίες είναι χρήσιμη κυρίως για να απαντηθούν ερωτήματα (ή ερευνητικές υποθέσεις) που σχετίζονται με την εγκυρότητα περιεχομένου και την εγκυρότητα βαθμολόγησης.

Για παράδειγμα, αν θελήσει να ελέγξει κανείς «πόση αντιπροσωπευτικότητα υπάρχει στους τύπους δοκιμασιών που αξιοποιήθηκαν σε ένα συγκεκριμένο τεστ ή σε όλα τα τεστ που έχει χρησιμοποιήσει ένα σύστημα πιστοποίησης σε σχέση με την τυπολογία δοκιμασιών που ορίζεται στις προδιαγραφές», θα χρειαστεί να κάνει κατηγοριοποίηση των δοκιμασιών που συναντά στα τεστ, καταμέτρηση των εμφανίσεων του κάθε τύπου και μετατροπή του αριθμού σε εκατοστιαία αναλογία. Έτσι θα καταλήξει σε έναν απλό *πίνακα μονής εισόδου* (single entry table), που φανερώνει αμέσως την αντιπροσωπευτικότητα των χρησιμοποιηθέντων τύπων, θεωρώντας τους δείγμα του συνόλου των τύπων δοκιμασιών που υιοθετεί το σύστημα. Στον πίνακα αυτό, *μέτρο κεντρικής τάσης* (measure of central tendency) θα θεωρηθεί η *επικρατούσα τιμή* (mode), η οποία δεν είναι τίποτε άλλο από τον τύπο που εμφανίστηκε περισσότερες φορές από κάθε άλλον και που, επομένως, υπερεκπροσωπείται. Οι μηδενικές εμφανίσεις χρειάζεται να τραβήξουν περισσότερο την

³⁸ Για ποσοτικό εσωτερικό και εξωτερικό έλεγχο ιδιαίτερα χρήσιμα είναι τα δημοσιεύματα van der Linden & Hambleton (1997) και Bachman (2004).

προσοχή αυτού που κάνει τον έλεγχο, γιατί μπορεί να οφείλονται σε διάφορους παράγοντες και ίσως να οδηγήσουν το σύστημα στην απόφαση αφαίρεσης των συγκεκριμένων τύπων από τις προδιαγραφές, αν η μη χρησιμοποίησή τους είναι εσκεμμένη και βάσιμη.

Κατά τον έλεγχο της εγκυρότητας βαθμολόγησης, τα δεδομένα που συλλέγονται είναι από την αρχή ποσοτικά –τουλάχιστον με βάση τον ορισμό των ποιοτικών και ποσοτικών δεδομένων που έχουμε δώσει σε αυτό το άρθρο.

Η τρίτη διαδικασία χρειάζεται περισσότερο κατά τον έλεγχο της εγκυρότητας περίπτωσης, όψης και συνεπειών.

Ένα παράδειγμα, σχετικό με την εγκυρότητα περίπτωσης, είναι η διερεύνηση του ερωτήματος «πόσο ενδιαφέρουσες βρίσκουν οι εξεταζόμενοι τις δοκιμασίες, δεδομένου του τύπου και του αριθμού τους;». Στην περίπτωση αυτή χρειάζεται συλλογή πληροφορίας, μέσω ερωτηματολογίου ή συνέντευξης, επί δείγματος εξεταζομένων, τη στιγμή που τα άτομα αποχωρούν από το χώρο των εξετάσεων ή και αργότερα.

Στο παραπάνω παράδειγμα, όπως είναι φανερό, η πληροφορία που συλλέγεται είναι ποιοτική, αλλά μετατρέπεται σε ποσότητες μέσω της αρίθμησης για να υποστεί στη συνέχεια στατιστική ανάλυση περιγραφικού τύπου, δηλαδή υπολογισμό της κεντρικής τάσης και της διασποράς (variation). Η συσχέτιση (correlation)³⁹ δεδομένων από την ίδια ομάδα (για παράδειγμα για να δούμε αν ένας τύπος δοκιμασίας αρέσει περισσότερο στους άντρες από ό,τι στις γυναίκες, ή αντίστροφα), ή και από διαφορετικές ομάδες υποκειμένων, εντάσσεται επίσης σε αυτή την τρίτη διαδικασία ελέγχου. Μόνο που εδώ η συσχέτιση δεν επιτρέπεται να ερμηνευτεί ως σχέση αιτίου αιτιατού, ούτε επιτρέπει να θεωρήσουμε ότι τα αποτελέσματα της συσχέτισης θα ήταν τα ίδια αν τα δεδομένα αφορούσαν στον/στους πληθυσμό/-ούς και όχι σε δείγμα/-ατα. Τέτοια ερμηνεία, αντίθετα, είναι επιτρεπτή στην τέταρτη διαδικασία που ορίσαμε.

Η τελευταία διαδικασία εφαρμόζεται κυρίως για τον έλεγχο της εγκυρότητας πρόβλεψης. Η εγκυρότητα πρόβλεψης, –μαζί με την *συντρέχουσα* ή *συνυπάρχουσα εγκυρότητα* (concurrent validity) που δεν περιλάβαμε στις βασικές διαστάσεις της εγκυρότητας– παρουσιάζει την ιδιαιτερότητα ότι απαιτεί την ύπαρξη ή τη συλλογή δεδομένων που δεν προέρχονται από το σύστημα πιστοποίησης. Αυτά τα άλλα δεδομένα θεωρούνται «κριτήριο», δηλαδή σημείο σύγκρισης.

Όπως έχει αναφερθεί, η βαθμολογία που προκύπτει από το γλωσσικό τεστ υποθέτουμε πως αποτελεί αξιόπιστη και δίκαιη ποσοτική έκφραση της γλωσσικής παραγωγής του εξεταζόμενου. Και η γλωσσική παραγωγή υποθέτουμε πως είναι εκδήλωση κυρίως των

³⁹ Ο όρος συσχέτιση χρησιμοποιείται μόνο στην περίπτωση σύγκρισης ποσοτικών δεδομένων. Ως υπερώνυμό της, δηλαδή ως όρο που να περιλαμβάνει αντιπαραβολή δεδομένων και ποιοτικού χαρακτήρα μπορούμε να θεωρήσουμε τη *σύγκριση* (comparison). Χρήσιμο για όποιον ενδιαφέρεται για τη σύγκριση στις κοινωνικές επιστήμες είναι το βιβλίο της C. Vigour (2005), όλο αφιερωμένο σε αυτό το θέμα.

ικανοτήτων και των γνώσεων που ο εξεταζόμενος διαθέτει, αλλά και των στάσεών του ως *επικοινωνητής* (communicator). Αν οι παραπάνω δύο υποθέσεις είναι αληθείς ή βάσιμες, τότε η επίδοσή του ως επικοινωνητής ή *χρήστης* (user) της γλώσσας στην οποία πιστοποιήθηκε –δηλαδή η επιτυχία με την οποία χρησιμοποιεί τη γλώσσα μετά από την εξέταση, σε φυσικές περιστάσεις επικοινωνίας (εργασιακός χώρος, ακαδημαϊκό περιβάλλον, περιήγηση σε χώρα όπου μιλιέται η γλώσσα, κοινωνικές συναναστροφές οπουδήποτε)– περιμένουμε να είναι αντίστοιχη του επιπέδου για το οποίο πιστοποιήθηκε και της βαθμολογίας που απέσπασε.

Το κριτήριο είναι ακριβώς το μέτρο της μελλοντικής επιτυχίας στη χρήση της γλώσσας εκτός εξετάσεων. Το πρόβλημα έγκειται στην εξασφάλιση αυτού του κριτηρίου, που μπορεί να αλλάζει κάθε φορά: άλλο κριτήριο θα πάρουμε αν θέλουμε να ελέγξουμε την εγκυρότητα πρόβλεψης ενός συστήματος που χορηγεί πιστοποιητικά για ακαδημαϊκή χρήση, άλλο για ένα σύστημα που πιστοποιεί τη γλωσσομάθεια ενηλίκων που βρίσκονται στην αγορά εργασίας, κ.τ.λ.

Η εγκυρότητα πρόβλεψης και η συντρέχουσα εγκυρότητα (που απαιτεί την ύπαρξη ενός κριτηρίου που είναι ταυτόχρονο: για παράδειγμα ένα σταθμισμένο και αποδεκτό ως έγκυρο τεστ χρησιμοποιείται την ίδια χρονική περίοδο για να συσχετίσουμε τα αποτελέσματά του με τα αποτελέσματα του υπό έλεγχο συστήματος πιστοποίησης), επειδή επιτρέπουν εξαγωγή συμπερασμάτων βάσει δεικτών συνάφειας, με όση ακρίβεια θέλουμε, θεωρούνται οι περισσότερο αντικειμενικά ελεγχόμενοι τύποι εγκυρότητας και συναντούν ιδιαίτερη προτίμηση στην Αμερική. Την τελευταία δεκαετία, ωστόσο, όλο και περισσότερο αμφισβητείται η ακρίβεια και η αντικειμενικότητα του ελέγχου τους, επειδή είναι δύσκολο να αποφασίσει κανείς τι ακριβώς θα θεωρηθεί κριτήριο και πώς ακριβώς θα κάνει μέτρηση του κριτηρίου αν τα δεδομένα δεν υπάρχουν ήδη.

Εν είδη επιλόγου πρέπει να σημειωθεί ότι η σύντομη αυτή εξέταση της εγκυρότητας και των τρόπων ελέγχου της δεν εξαντλεί καθόλου το θέμα. Στόχος ήταν αφενός η ανάδειξη της συνθετότητας της έννοιας εγκυρότητα και αφετέρου η εστίαση σε κάποια χαρακτηριστικά της και η σκίαση άλλων προς διευκόλυνση των ερευνητών-αναγνωστών που ασχολούνται ή πρόκειται να ασχοληθούν με το θέμα της εγκυρότητας συστημάτων πιστοποίησης της γλωσσομάθειας.

Βιβλιογραφικές παραπομπές

- Anastasi, A. (1986). Evolving Concepts of Test Validation. In *Annual Review of Psychology*, 37: 1-15.
- Anastasi, A. (1988⁶). *Psychological Testing*. New York: Macmillan.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Berthier, N. (2006³). *Les techniques d'enquête en sciences sociales*. Paris: Armand Colin.
- Blalock, H. M. (1984). *Statistica per la ricerca sociale*. Bologna: Il Mulino (Μετάφραση του V. Mortara από την αγγλική, όπου δημοσιεύτηκε το 1979 από τον οίκο McGraw-Hill, N.Y.).
- Braun, H. (2000). A Post-modern View of the Problem of Language Assessment. In Kunnan, A. J. (Ed.). *Fairness and Validation in Language Assessment*. Cambridge: Cambridge University Press.
- Chapelle, C. A. (1999). Validity in Language Assessment. In *Annual Review of Applied Linguistics* 19: 254-272. Cambridge: Cambridge University Press.
- Common European Framework of Reference for Languages: Learning, teaching, assessment* (2001). Cambridge: Cambridge University Press (Το βιβλίο είναι διαθέσιμο στο http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf. Η απόδοση του κειμένου στη γαλλική γλώσσα είναι προσβάσιμη στη διεύθυνση http://www.coe.int/T/DG4/Linguistic/Source/Framework_FR.pdf Τέλος, για την απόδοση στην ελληνική ο αναγνώστης παραπέμπεται στη διεύθυνση <http://www.komvos.edu.gr/Common%20European%20Framework.pdf>).
- Cronbach, L. J., Meehl P. E. (1955). Construct validity in psychological tests. In *Psychological Bulletin*, 52: 281-302 (Όλο το άρθρο είναι διαθέσιμο και στη διεύθυνση <http://psychclassics.yorku.ca/Cronbach/construct.htm>).
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara T. (1999). *Dictionary of Language Testing*. Cambridge: Cambridge University Press.
- De Ketele, J.-M., Roegiers, X. (1991). *Méthodologie du recueil d'informations*. Bruxelles: De Boeck Université.
- De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris: Presses Universitaires de France.
- Dorvil, H. (sous la direction de) (2007). *Problèmes sociaux, Tome III: Théories et méthodologies de la recherche*. Québec: Presses de l'Université du Québec.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests*. New York: McGraw-Hill.
- Langouet, G., Porlier, J.-C. (1991⁴). *Mesure et statistique en milieu éducatif*. Paris: ESF.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.) *Educational Measurement*, 3rd edition (13-103). New York: Macmillan.
- Quivy, R. & Van Campenhoudt, L. (1988). *Manuel de recherché en sciences sociales*. Paris: Bordas.

- Trochim, W. M. K. (2006). Introduction to Validity. [Social Research Methods](#).
- Van der Linden, W., Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Van der Maren, J.-M. (1996). *Méthodes de recherche pour l'éducation*. Montréal: Presses de l'Université de Montréal & Paris-Bruxelles: De Boeck Université.
- Verma, G. K., Mallick, K. (2004). *Εκπαιδευτική Έρευνα: Θεωρητικές προσεγγίσεις και τεχνικές*. Αθήνα: Τυπωθήτω - Γ. Δαρδάνος (Μετάφραση της Ε. Γρίβα από την αγγλική, στην οποία δημοσιεύτηκε το 1999).
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-based Approach*. Houndmills, Hampshire: Palgrave Macmillan.
- Γεωργούσης, Π. (1999). *Η μέτρηση και η αξιολόγηση της επίδοσης των μαθητών*. Αθήνα.
- Τσαγγαρή, Ντ. (2011). Επιπτώσεις στην εκπαίδευση και τη μάθηση από τις εξετάσεις πιστοποίησης. [RCEL-Research centre for english language](#).
- Τσάντας, Ν., Μωυσιάδης, Χ., Μπαγιάτης, Ν., Χατζηπαντελής, Θ. (1999). *Ανάλυση δεδομένων με τη βοήθεια στατιστικών πακέτων*. Θεσσαλονίκη: Ζήτη.
- Τσομπάνογλου, Μ. (2007). *Πρότυπο αξιολόγησης της εγκυρότητας δομής συστημάτων πιστοποίησης γλωσσομάθειας. Εφαρμογή του στο Κρατικό Πιστοποιητικό Γλωσσομάθειας του επιπέδου B2 της αγγλικής γλώσσας*. Αδημοσίευτη διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- Τσομπάνογλου, Α. (2010²). *Μεθοδολογία της επιστημονικής έρευνας και εφαρμογές της στην αξιολόγηση της γλωσσικής κατάρτισης*. Θεσσαλονίκη: Ζήτη.