

Θεωρίες για τις μετρήσεις στην Εκπαιδευτική Έρευνα

Θάνος Βέρδης

Η κλασική θεωρία των μετρήσεων

Η ΕΝΝΟΙΑ ΤΟΥ ΠΡΑΓΜΑΤΙΚΟΥ ΣΚΟΡ

Η πρώτη θεωρία με την οποία οι ψυχομέτρες προσέγγισαν τις λανθάνουσες μεταβλητές και ιδιαίτερα την έννοια της νοημοσύνης ήταν η Θεωρία του Πραγματικού Σκορ ή αλλιώς η Κλασική Θεωρία των μετρήσεων. Η θεωρία αυτή αναπτύχθηκε ως εξέλιξη των πρώτων τεστ νοημοσύνης. Το «πραγματικό σκορ» στην κλασική θεωρία των μετρήσεων είναι ένας αριθμός και συγκεκριμένα ο αριθμητικός μέσος όρος των παρατηρούμενων σκορ που θα βρίσκαμε αν επαναλαμβάναμε το ίδιο τεστ πολλές φορές στους ίδιους μαθητές. Ας υποθέσουμε ότι κάναμε στους μαθητές μας ένα τεστ ορθογραφικής ικανότητας. Τη μία ημέρα μπορεί οι μαθητές μας να ήταν κουρασμένοι και να μην απέδωσαν καλά ή αντίθετα την επόμενη ημέρα να απέδωσαν καλά λόγω τύχης. Έτσι, αν το καθημερινό παρατηρούμενο σκορ του μαθητή είναι x και η πραγματική γνώση ορθογραφίας είναι τ , η καθημερινή διαφορά μεταξύ τους είναι E και ισχύει ότι $x = \tau + e$. Η Κλασική Θεωρία των Τεστ είναι η θεωρία του αληθινού σκορ και των σφαλμάτων. Σύμφωνα με αυτήν, κάθε μέτρηση στις επιστήμες του ανθρώπου εμπεριέχει ένα αληθινό σκορ και ένα σφάλμα, το οποίο μπορεί να είναι τυχαίο ή και συστηματικό. Οι στατιστικολόγοι έχουν αποδείξει ότι το σφάλμα είναι το ίδιο ανεξάρτητα αν εφαρμόζουμε το τεστ μία φορά σε ένα πλήθος συμμετεχόντων ή πολλές φορές σε έναν και μόνο συμμετέχοντα (βλ. Allen, M.J., & Yen, 2002). Η διακύμανση των μετρήσεων είναι στην ουσία η διακύμανση του πραγματικού σκορ συν τη διακύμανση του σφάλματος. Η αξιοπιστία ενός τεστ, η οποία συμβολίζεται με R , αποδεικνύεται ότι είναι $R = 1 - \text{var}(E) / \text{var}(X)$ και ότι $\text{var}(T) = \text{var}(X) \times R$. Άρα, αν γνωρίζουμε την αξιοπιστία και τη διακύμανση στο παρατηρούμενο σκορ μπορούμε να προσεγγίσουμε το πραγματικό σκορ. Όσο περισσότερα ερωτήματα έχει ένα τεστ τόσο μεγαλώνει η αξιοπιστία του και τόσο περισσότερο προσεγγίζουμε το πραγματικό σκορ.

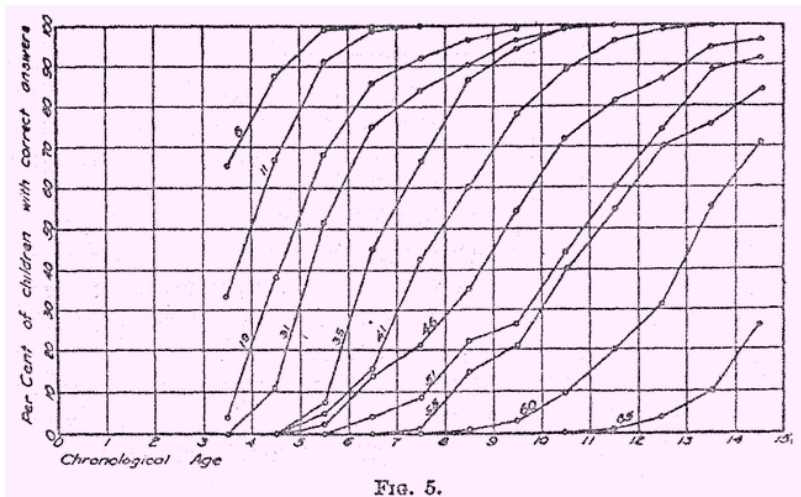


Fig. 5.

Εικόνα 1. Από το 1925 υπήρχε επίγνωση ότι τα ερωτήματα ενός τεστ δεν έχουν τον ίδιο βαθμό δυσκολίας. Ποσοστά ορθών ανταποκρίσεων στις δοκιμασίες του τεστ των Binet-Simon σε παιδιά διαφορετικών ηλικιών (πηγή Thurstone, 1925: 444). Οι αριθμοί πάνω στις γραμμές αναφέρονται στον αριθμό των δοκιμασιών.

Κάθε ερώτημα σε ένα τεστ μπορεί σύμφωνα με την Κλασική Θεωρία να θεωρηθεί εύκολο ή δύσκολο, ανάλογα με το πόσο συμμετέχοντες το απάντησαν σωστά. Με τον τρόπο αυτόν τα ερωτήματα διαφέρουν αρχικά ως προς τη χρησιμότητά τους. Όπως γίνεται εύκολα αντιληπτό, αυτά που απαντιούνται απ' όλους και αυτά που δεν απαντώνται από κανέναν δεν είναι πολύ χρήσιμα για τους ερευνητές και θα μπορούσαν να παραλειφθούν από το τεστ. Στην Κλασική Θεωρία επίσης, κάθε ερώτημα χαρακτηρίζεται από τη δυνατότητα διάκρισης που έχει, ανάλογα με το πόσο καλά διαφοροποιεί τους ικανούς και

τους λιγότερο ικανούς μαθητές. Τη δυνατότητα διάκρισης ενός ερωτήματος την υπολογίζουμε στην Κλασική Θεωρία μέσω της διαφοράς στην παρατηρούμενη βαθμολογία που έχει τριάντα περίπου τοις εκατό των υψηλότερων βαθμολογιών από το τριάντα τοις εκατό των χαμηλότερων βαθμολογιών. Αυτή η δυνατότητα διάκρισης συνδέεται με τον συντελεστή συσχέτισης κάθε ερωτήματος με όλα τα άλλα ερωτήματα του τεστ. Όσο πιο πολύ σχετίζεται ένα θέμα με όλα τα άλλα θέματα τόσο μεγαλύτερη αξιοπιστία θεωρείται ότι διαθέτει ένα τεστ. Μέσω της Κλασικής Θεωρίας λοιπόν, οι ερευνητές μπορούν να υπολογίσουν το πραγματικό σκορ, καθώς και τις ψυχομετρικές ιδιότητες κάθε μεμονωμένου ερωτήματος ενός τεστ.

Η Κλασική Θεωρία όμως έχει ορισμένα προβλήματα. Κατ' αρχάς προϋποθέτει ότι όλες οι ερωτήσεις σε ένα τεστ μετρούν το πραγματικό σκορ με τον ίδιο ακριβώς τρόπο. Για παράδειγμα, αν έχουμε ένα τεστ ή ένα ερωτηματολόγιο με 10 ερωτήσεις τύπου Likert -ας υποθέσουμε 1= «λίγο», 2= «αρκετά», 3= «πολύ» και 4= «πάρα πολύ»- υπάρχουν πάνω από ένα εκατομμύριο και σαράντα οκτώ χιλιάδες πιθανοί συνδυασμοί απαντήσεων. Μέσα σε αυτούς τους συνδυασμούς υπάρχουν πάρα πολλοί συνδυασμοί με τους οποίους μπορεί να πάρει κανείς ένα σκορ ως πούμε ίσο με 20. Η κλασική θεωρία υποστηρίζει ότι όλοι οι συνδυασμοί που δίνουν σκορ 20 έχουν την ίδια ουσιαστική βαρύτητα στη μέτρηση της θεωρητικής κατασκευής. Αυτό, όμως, δεν είναι αλήθεια διότι κάθε ερώτημα συνδέεται με διαφορετικό τρόπο με τη θεωρητική κατασκευή που θέλουμε να μετρήσουμε. Άλλο είναι, για παράδειγμα, να απαντήσει κάποιος «πολύ» στην ερώτηση «κατά πόσο συμφωνείτε με την πρόσφατη θέσπιση του πολιτικού γάμου» και άλλο είναι να απαντήσει κανείς «πολύ» στην ερώτηση «πόσο πιθανό είναι να παντρευτείτε εσείς προσωπικά με πολιτικό γάμο». Αν και θεωρητικά υπάρχει δυνατότητα και στατιστικοί τρόποι ώστε τα ερωτήματα ενός τεστ να υπολογίζονται με διαφορετικούς συντελεστές βαρύτητας το καθέναν (π.χ. μέσω Ανάλυσης Παραγόντων ή μέσω του συντελεστή συνάφειας ενός ερωτήματος με τα υπόλοιπα ερωτήματα κλπ.), η διαφοροποιημένη βαρύτητα δεν επηρεάζει δομικά το τεστ. Μπορεί, για παράδειγμα, μέσω διαφορετικών συντελεστών βαρύτητας να πάρουμε διαφορετικά τελικά σκορ για δύο εξεταζόμενους που χωρίς τους συντελεστές θα είχαν

δομήσει το ίδιο σκορ διαφορετικά, αλλά η εσωτερική εγκυρότητα και αξιοπιστία των αρχικών ερωτημάτων δεν θα αλλάζει.

Ένα άλλο σοβαρό ζήτημα σε σχέση με την κλασική θεωρία είναι ότι οι κλίμακες μέτρησης της θεωρητικής κατασκευής πρέπει να είναι τουλάχιστον ίσων διαστημάτων και όχι τακτικών τιμών ή πολύ περισσότερο διχοτομικές (βλ. Linacre, 2005). Κλίμακες ίσων διαστημάτων είναι αυτές στις οποίες η κλίμακα μέτρησης έχει ίσα διαστήματα από το ένα σημείο στο άλλο. Κλίμακα ίσων διαστημάτων είναι, για παράδειγμα, μια κλίμακα που θα μετρούσε την ηλικία ενός μαθητή σε μήνες ή μια κλίμακα που θα μετρούσε τον αριθμό των λέξεων ενός κειμένου. Κλίμακες τακτικών τιμών ή άνισων διαστημάτων από την άλλη μεριά είναι κλίμακες στις οποίες τα διαστήματα από μία τιμή στην άλλη δεν είναι ίσα. Τέτοιες κλίμακες έχουμε για παράδειγμα κατά την καταγραφή θετικής ή αρνητικής άποψης σε μια έρευνα διερεύνησης της κοινής γνώμης, στις οποίες οι πιθανές απαντήσεις είναι του τύπου «ελάχιστα ευχαριστημένος», «αρκετά ευχαριστημένος», «πολύ ευχαριστημένος» και «πάρα πολύ ευχαριστημένος». Η απόσταση μεταξύ του «ελάχιστα ευχαριστημένος» και του «αρκετά ευχαριστημένος» δεν είναι ίση με την απόσταση μεταξύ του «πολύ ευχαριστημένος» και του «πάρα πολύ ευχαριστημένος». Διχοτομική κλίμακα είναι αυτή στην οποία οι πιθανές απαντήσεις είναι δύο και σε αντίθετες κατευθύνσεις μεταξύ τους: το «ναι» και το «όχι» με βάση το αν πέτυχε κάποιος σε ένα τεστ, το «σωστό» και το «λάθος» στη γραφή μιας λέξης, το «ευχαριστημένος» και το «δυσανεστημένος» σε ένα ερωτηματολόγιο κλπ.

Η ανάλυση δεδομένων τακτικών τιμών επηρεάζεται από το εύρος των διαθέσιμων επιλογών. Πράγματι, τα αποτελέσματα της ανάλυσης είναι πολύ διαφορετικά αν η κλίμακα που χρησιμοποιούμε έχει λίγες ή περισσότερες πιθανές κατηγορίες. Άλλη συσχέτιση έχουν για παράδειγμα κλίμακες με τέσσερα άνισα διαστήματα και άλλη συσχέτιση έχουν κλίμακες με οκτώ άνισα διαστήματα. Άλλη συμπεριφορά έχουν οι κλίμακες στις οποίες υπάρχει κάποια μεσαία ή «ουδέτερη» κατηγορία (π.χ. «ούτε ευχαριστημένος, ούτε δυσανεστημένος») και άλλη συμπεριφορά έχουν οι κλίμακες οι οποίες κατευθύνουν αναγκαστικά τον ανταποκρινόμενο είτε προς την πλευρά μιας γενικότερης ικανοποίησης είτε προς την πλευρά μιας γενικότερης μη ικανοποίησης χωρίς τη δυνατότητα κάποιας ουδέτερης ή μεσαίας επιλογής. Οι κλίμακες διαφέρουν επίσης ως προς την ύπαρξη κάποιας «μηδενική» κατηγορίας όπως θα ήταν το «καθόλου ευχαριστημένος», το οποίο διαφέρει σε τεράστιο βαθμό από το «λίγο ευχαριστημένος» και από το υπόλοιπο μέρος της κλίμακας.

ΛΑΝΘΑΝΟΥΣΕΣ ΜΕΤΑΒΛΗΤΕΣ ΚΑΙ ΑΝΑΛΥΣΗ ΠΑΡΑΓΟΝΤΩΝ

Στο πλαίσιο της κλασικής θεωρίας των μετρήσεων εντάσσονται και κάποιες τεχνικές ανάλυσης με τις οποίες οι ερευνητές μετρούν λανθάνουσες μεταβλητές, δηλαδή μεταβλητές που δεν παρατηρούνται ευθέως. Τέτοιες μεταβλητές έχουν να κάνουν, για παράδειγμα, με τις στάσεις και τις προσδοκίες των ανθρώπων, τις γνώσεις τους, τις συμπεριφορές τους, τις δεξιότητές τους. Διάφοροι μελετητές έχουν εκφράσει διαφορετικές απόψεις για το πώς τελικά θα έπρεπε να εννοούνται οι λανθάνουσες μεταβλητές στην ψυχολογική αλλά και την παιδαγωγική και την κοινωνική έρευνα. Ο Bollen (2002) έχει συγκεντρώσει κριτικά πολλές κατηγορίες ορισμών πριν δώσει τελικά τον δικό του, σύμφωνα με τον οποίο: «μια τυχαία ή μη τυχαία μεταβλητή ονομάζεται “λανθάνουσα” όταν κάποιες από τις παρατηρήσεις της δεν πραγματώνονται σε ένα ορισμένο δείγμα» (Bollen, ό. π.: 612).

Βασική ιδέα στην κλασική θεωρία της μέτρησης είναι ότι οι παρατηρήσεις συνδέονται με ένα σύνολο λανθανουσών μεταβλητών, οι οποίες έχουν τη μορφή κοινών παραγόντων. Επιδίωξη

της ανάλυσης είναι να εκφραστούν οι παρατηρήσεις (έστω x_1 ως x_p) με τους κοινούς αυτούς παράγοντες f_1 ως f_q αφού φυσικά αυτοί πρώτα αναγνωριστούν τόσο ως προς τον αριθμό τους, όσο και ως προς τη φύση τους. Αν φυσικά υπάρχει μια εκ των προτέρων θεωρία για τον αριθμό και τη φύση αυτών των παραγόντων τότε η θεωρία αυτή τίθεται υπό δοκιμασία με βάση τα δεδομένα που συλλέγονται. Στις παρακάτω εκφράσεις τα λ είναι οι συντελεστές των παραγόντων f ενώ τα u είναι το μέρος των x που δεν εξηγείται από τους κοινούς παράγοντες. Τα μ_i είναι κάποιες ποσότητες για την περίπτωση που τα f είναι μηδέν, ενώ τα λ_i είναι οι συντελεστές των f . Οι παρακάτω εξισώσεις θυμίζουν γραμμική παλινδρόμηση, μόνο που τώρα τα f δεν είναι «ανεξάρτητες μεταβλητές». Στην ανάλυση παραγόντων τα ίδια τα δεδομένα θα ορίσουν τον αριθμό και τη φύση των f :

$$\begin{aligned} x_1 &= \mu_1 + \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1q}f_q + u_1 \\ x_2 &= \mu_2 + \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2q}f_q + u_2 \\ &\vdots \\ x_p &= \mu_p + \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pq}f_q + u_p \end{aligned}$$

Οι παραπάνω σχέσεις γράφονται με μορφή πινάκων ως $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{u}$. Τα $x_1 \dots x_p$ αφορούν p παρατηρήσεις σε έναν μόνο μέλος του δείγματος. Για ένα δείγμα μεγέθους n η προηγούμενη ισότητα γράφεται $\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i + \mathbf{u}_i$ με $i = 1, 2, \dots, n$.

Στην πιο πάνω έκφραση τα $\boldsymbol{\mu}$ είναι οι τιμές όταν οι παράγοντες η προσδοκία των παραγόντων είναι μηδέν δηλαδή $E(\mathbf{x})$ όταν $f_1 = f_2 = \dots = f_q = 0$. Το μη εξηγούμενο μέρος, δηλαδή τα u , είναι ανεξάρτητο από το κοινό μέρος, δηλαδή τα f δηλαδή η συνδιακύμανση $\text{cov}(u, f) = 0$. Ο μέσος όρος των u είναι μηδέν $E(\mathbf{u}) = 0$. Τέλος, ο πίνακας διακυμάνσεων-συνδιακυμάνσεων των u είναι διαγώνιος $\text{cov}(\mathbf{u}) = \boldsymbol{\Psi}$ με τη διακύμανση στη διαγώνιο και τις συνδιακυμάνσεις μηδέν. Όταν λέμε ότι ένας πίνακας είναι «διαγώνιος» εννοούμε ότι δεν υπάρχει σχέση μεταξύ των u . Για δύο μόνο u η κατανομή θα έμοιαζε με το αριστερό τμήμα στο. **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε., σελ. Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης..** Για περισσότερα u η κατανομή θα είχε περισσότερες διαστάσεις.

Έχουμε λοιπόν ότι τα u και οι παράγοντες f είναι ανεξάρτητοι μεταξύ τους. Έχουμε επίσης ότι ο $\boldsymbol{\Psi}$ είναι διαγώνιος $\text{cov}(\mathbf{u}) = \boldsymbol{\Psi}$, καθώς και ότι $\text{cov}(\mathbf{f}) = \mathbf{I}$ (όπου \mathbf{I} ο μοναδιαίος πίνακας). Με όλα αυτά υπόψη ο πίνακας διακυμάνσεων-συνδιακυμάνσεων των x είναι $\boldsymbol{\Sigma}$ και είναι:

$$\text{cov}(x) = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$$

Αντιστρόφως, αν ο πίνακας $\boldsymbol{\Sigma}$ μπορεί να γραφτεί ως $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$ τότε υπάρχουν κοινοί παράγοντες f .

Η γενικότερη ιδέα της εύρεσης ενός αριθμού κοινών λανθανουσών μεταβλητών οι οποίες από κοινού «εξηγούν» μέρος της διακύμανσης στα x αποτελεί την ουσία της ανάλυσης παραγόντων και είναι μια βασική κατάκτηση της κλασικής θεωρίας των μετρήσεων. Η διακύμανση που από κοινού εξηγούν οι κοινοί παράγοντες συμβολίζεται με h^2 .

Η συνολική διακύμανση στα x είναι $\sum_{j=1}^k \lambda_{ij}^2 + u_{ij}$. Η ποσότητα $h_i^2 = \sum_{j=1}^k \lambda_{ij}^2$ ονομάζεται “communality” και όπως είπαμε είναι η διακύμανση που επεξηγείται από τους κοινούς παράγοντες. Οι κοινοί αυτοί παράγοντες θα μπορούσαν θεωρητικά να εξηγήσουν όλη τη διακύμανση στα x αν το πλήθος τους ήταν ίσο με το πλήθος των x . Στη περίπτωση αυτή, όμως, δεν θα είχε νόημα η ανάλυση παραγόντων διότι θα επρόκειτο στην ουσία για μια διαφορετική έκφραση των ίδιων μεταβλητών. Στην ουσία αυτό που θέλουμε είναι να μειώνουμε τον αριθμό των παρατηρούμενων μεταβλητών σε έναν χώρο λιγότερων διαστάσεων. Άρα στην ανάλυση παραγόντων τα f είναι πάντα λιγότερα από τα x .

Οι συντελεστές των f μπορεί να είναι διαφορετικοί και θεωρητικά υπάρχουν πολλοί ισοδύναμοι πίνακες Λ . Σε μια διερευνητική ανάλυση παραγόντων είναι στο χέρι των ερευνητών να επιλέξουν κατάλληλους συντελεστές των f θέτοντας περιορισμούς στις διακυμάνσεις μέσω κατάλληλων αλγόριθμων όπως ο Varimax ή ο Oblimim. Για να υπολογιστούν τα λ (ας υποθέσουμε ότι ο αριθμός των f έχει ήδη καθοριστεί) υπάρχουν δύο γνωστές μέθοδοι στη βιβλιογραφία: η μέθοδος της μέγιστης πιθανοφάνειας και η μέθοδος των κύριων παραγόντων. Στην πρώτη, η οποία είναι μέθοδος που διευκολύνθηκε από τους σημερινούς γρήγορους υπολογιστές, μεγιστοποιούμε τη πιθανοφάνεια του Λ ως προς τον Ψ . Διάφοροι στατιστικοί δίνουν την ποσότητα που μεγιστοποιούμε με τη μέθοδο αυτή (βλ. Kreiner & Christensen, 2014· Stoica & Jansson, 2009· Thomson, 1980). Ως λογαριθμική πιθανοφάνεια είναι:

$$\log L = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}).$$

Τέλος στην εκπαιδευτική έρευνα χρησιμοποιείται πολλές φορές εναλλακτικά προς την ανάλυση παραγόντων η ανάλυση κυρίων συνιστωσών (βλ. Kellow, 2006· Smith, 2002). Η ανάλυση κυρίων συνιστωσών, όμως, δεν είναι ανάλυση παραγόντων αλλά, όπως λέει το όνομά της, συνιστωσών, δηλαδή ένας γραμμικός μαθηματικός μετασχηματισμός των δεδομένων σε έναν μικρότερο αριθμό διαστάσεων. Όταν δε οι συντελεστές των συνιστωσών αλλάζουν ώστε αυτές να αποκτήσουν ουσιαστικότερη σχέση με τα αρχικά δεδομένα, τότε οι συνιστώσες δεν είναι «κύριες».

Παρόλα αυτά, η δομή που θέτει η ανάλυση κυρίων συνιστωσών στα δεδομένα μοιάζει με τη δομή που θέτει στα δεδομένα η ανάλυση παραγόντων. Η δομή των k κυρίων συνιστωσών είναι $\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{\Gamma} \mathbf{k}_i + \mathbf{e}_i$ όπου ο πίνακας $\mathbf{\Gamma}$ είναι ο πίνακας των συντελεστών και \mathbf{e}_i τα υπόλοιπα. Η διαφορά όμως της ανάλυσης παραγόντων από την ανάλυση κύριων συνιστωσών είναι ότι στην ανάλυση παραγόντων ο πίνακας $\text{cov}(\mathbf{u})$ είναι διαγώνιος, ενώ στην ανάλυση κυρίων συνιστωσών ο αντίστοιχος πίνακας $\text{cov}(\mathbf{e}_i)$ δεν είναι διαγώνιος. Έτσι οι σκοποί της ανάλυσης συνιστωσών και της ανάλυσης παραγόντων διαφέρουν ελαφρώς. Στην ανάλυση συνιστωσών έχουμε τη βέλτιστη δυνατή εξερεύνηση των διακυμάνσεων των x_i ενώ στην ανάλυση παραγόντων εξερευνούμε τις συνάψεις και τις συνδιακυμάνσεις. Στην πράξη η ανάλυση παραγόντων και η ανάλυση κύριων συνιστωσών δίνουν αποτελέσματα που δεν διαφέρουν εντυπωσιακά, ιδιαίτερα αν οι διακυμάνσεις είναι μικρές. Στη θεωρητική περίπτωση που $e = u = 0$ τα αποτελέσματά της ανάλυσης παραγόντων και της ανάλυσης κυρίων συνιστωσών θα ταυτίζονταν. Κάτι τέτοιο όμως

δεν είναι ρεαλιστικό. Και οι δύο τεχνικές, πάντως, επειδή βασίζονται σε διακυμάνσεις ή συνάψεις, απαιτούν κλίμακες ίσων διαστημάτων, αν και τελευταία έχουν αναπτυχθεί τεχνικές για αντίστοιχες αναλύσεις με τακτικές τιμές.

Η νέα θεωρία των μετρήσεων

Η ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ GEORG RASCH

Τα τελευταία χρόνια μια νέα θεωρία έχει προταθεί για τη ψυχολογική μέτρηση, η οποία υπόσχεται ότι ξεπερνάει πολλούς από τους περιορισμούς της κλασικής θεωρίας που είδαμε στις προηγούμενες παραγράφους. Η θεωρία αυτή είναι η Item Response Theory (IRT), δηλαδή η Θεωρία Ανάλυσης Αποκρίσεων ή σε ελεύθερη απόδοση η Νέα Ψυχομετρική Θεωρία. Οι απαρχές της θεωρίας αυτής ανιχνεύονται στο έτος 1943 στη δουλειά του Σκωτσέζου στατιστικού Derrick Lawley, ο οποίος εργαζόταν κοντά στον παιδαγωγό Godfrey Thomson σε ένα γραφείο στο Moorey House του Πανεπιστημίου του Εδιμβούργου. Βασική φιλοσοφία της εν λόγω θεωρίας είναι ότι οι μετρήσεις των διαφόρων χαρακτηριστικών των ανθρώπων εξαρτώνται τόσο από αυτά που θέλουμε να μετρήσουμε, όσο και από τα χαρακτηριστικά των θεμάτων των τεστ με τα οποία γίνεται η μέτρηση. Έτσι, με τη νέα θεωρία μετριέται ταυτοχρόνως το υποκείμενο και το εργαλείο της μέτρησης. Αυτή τη θεωρία χρησιμοποίησαν οι Σιδεριδής κ.ά. (2008) στην έρευνα που θα παρουσιάσουμε πιο κάτω.

Μεγάλη ώθηση στη νέα θεωρία των μετρήσεων έδωσε ο Δανός μαθηματικός Georg Rasch, ο οποίος πρότεινε μοντέλα για διχοτομικές απαντήσεις τύπου «σωστό» - «λάθος». Τα εν λόγω μοντέλα πήραν το όνομά του, όπως θα δούμε παρακάτω σε αυτό το κεφάλαιο. Σημαντική για τον τύπο αυτόν της μέτρησης ήταν επίσης και η συμβολή της Ιαπωνίδας μαθηματικού Fumiko Samejima, η οποία μελέτησε αντίστοιχες μεθόδους για περιπτώσεις που τα αποτελέσματα των τεστ δεν έχουν μόνο δύο αλλά περισσότερες από δύο κατηγορίες (Samejima, 1969). Οι Darrell Bock και David Thissen από το Πανεπιστήμιο του Σικάγου αλλά και άλλοι ερευνητές, όπως ο Αυστριακός David Andrich, εργάστηκαν ανεξάρτητα στην εν λόγω θεωρία και επέκτειναν την ιδέα της Samejima ως προς την τελειοποίηση κατάλληλων αλγορίθμων για τον υπολογισμό μοντέλων της θεωρίας αυτής για περιπτώσεις τεστ με πολλές κατηγορίες, ενώ Gerhard Fischer



Εικόνα 2 Ο Δανός Μαθηματικός Georg Rasch (αριστερά) με τον φυσικό Benjamin Wright, πρωτοπόρο στη χρήση των υπολογιστών για ψυχομετρικές αναλύσεις.

επέκτεινε προς την ίδια κατεύθυνση τα μοντέλα τύπου Rasch. Στην διάδοση της Θεωρίας των Αποκρίσεων σημαντική υπήρξε και η συμβολή του Αμερικανού φυσικού Benjamin Wright, ο οποίος ήταν από τους πρώτους που χρησιμοποίησαν ηλεκτρονικούς υπολογιστές με λυχνίες και με την τεχνολογία του 1959 για τις ανάγκες υπολογισμού των παραμέτρων στην ανάλυση δεδομένων μέσω Θεωρίας των Αποκρίσεων. Ο Wright συνεχίζει μέχρι σήμερα με τις σύγχρονες υπολογιστικές δυνατότητες να εργάζεται με τη Item Response Theory και να διαδίδει το λογισμικό Winsteps για αναλύσεις κατά Rasch. Έτσι λοιπόν έχουμε τρία μοντέλα Θεωρίας Αποκρίσεων, ανάλογα με τους πόσους παράγοντες χειρίζονται: (α) μοντέλα με ένα

παράγοντα (ή μοντέλα Rasch), (β) μοντέλα με δύο παράγοντες και (γ) μοντέλα με τρεις παράγοντες, αν και πολλοί υπολογίζουν τα μοντέλα Rasch ως μια κατηγορία από μόνη της, η οποία διαφέρει από τα μοντέλα με έναν παράγοντα. Στις επόμενες παραγράφους θα αναλύσουμε τα βασικά σημεία της Θεωρίας των Αποκρίσεων και ιδιαίτερα της ανάλυσης κατά Rasch. Στην Ελλάδα η θεωρία αυτή αναπτύχθηκε κατ' αρχάς στην ψυχολογική έρευνα με βάση τις εργασίες του καθηγητή του Πανεπιστημίου του Harvard Γιώργου Σιδερίδη (Sideridis, 2011). Ο πρώτος Έλληνας παιδαγωγός (όχι ψυχολόγος) που έγραψε για τη θεωρία αυτή ήταν ο καθηγητής του Τμήματος Φιλοσοφίας – Παιδαγωγικής – Ψυχολογίας του Πανεπιστημίου Αθηνών Μιχάλης Κασσωτάκης στο βιβλίο του *Η Αξιολόγηση της Επίδοσης των Μαθητών* (Κασσωτάκης, 2013).

Η Θεωρία των Αποκρίσεων όπως και η Κλασική Θεωρία δέχεται ότι υπάρχει μια λανθάνουσα μεταβλητή η οποία μετρείται με τα ερωτήματα του τεστ (τα items). Αυτή η μεταβλητή συμβολίζεται στην Θεωρία των Αποκρίσεων με θ . Δέχεται επίσης η Θεωρία των Αποκρίσεων ότι κάθε ερώτημα έχει και έναν βαθμό δυσκολίας, ο οποίος επηρεάζει τη σωστή απάντηση σε καθένα από αυτά. Η δυσκολία του ερωτήματος συμβολίζεται με δ . Η πιθανότητα να απαντήσει κάποιος σωστά σε ένα θέμα εξαρτάται από την ικανότητά αλλά και την δυσκολία του θέματος. Ισχύει δηλαδή η εξής γραμμική λογική: Πιθανότητα ορθής απάντησης = ικανότητα του υποψηφίου – δυσκολία του θέματος.

Με την πιο πάνω λογική αν ένας μαθητής μέτριας ικανότητας η καταπιαστεί με ένα θέμα μέτριας δυσκολίας i τότε $\theta_n - \delta_i = 0$. Στην περίπτωση αυτή ο μαθητής έχει ίσες πιθανότητες να απαντήσει ή να μην απαντήσει στο ερώτημα και άρα η πιθανότητα σωστής απάντησης είναι 0,5. Αν η ικανότητα του μαθητή είναι μεγαλύτερη από τη δυσκολία του ερωτήματος τότε $\theta_n - \delta_i > 0$ και ο μαθητής έχει πιθανότητα να απαντήσει στο ερώτημα μεγαλύτερη από 0,5. Αν η ικανότητα του μαθητή είναι μικρότερη από τη δυσκολία του ερωτήματος τότε $\theta_n - \delta_i < 0$ και η πιθανότητα ορθής απάντησης είναι μικρότερη από 0,5. Για παράδειγμα, ας υποθέσουμε ότι ρωτήσαμε έναν αριθμό μαθητών να μας γράψουν τις λέξεις «από», «αυτοκίνητο», «διαδίκτυο», «κωλυσιεργία», και «αντενδείκνυται». Η πιθανότητα να απαντήσει κάποιος ορθά σε κάποια από αυτές τις λέξεις υπολογίζεται με βάση (α) το πώς έχουν απαντήσει οι υπόλοιποι εξεταζόμενοι στην ίδια λέξη και (β) πώς έχει απαντήσει ο ίδιος εξεταζόμενος στις υπόλοιπες ερωτήσεις. Η σωστή ή λάθος απάντηση εξαρτάται από την ορθογραφική του ικανότητα μείον την δυσκολία της λέξης. Ας υποθέσουμε ότι ένας εξεταζόμενος μπορεί να γράψει ορθογραφημένα μια λέξη με πιθανότητα 40 τοις εκατό ή 0,40. Η πιθανότητα να μην γράψει ορθογραφημένα τη λέξη είναι το υπόλοιπο 60 τοις εκατό ή 0,60, αφού $1 - 0,40 = 0,60$. Αυτή η τελευταία είναι «συμπληρωματική πιθανότητα» του 0,40. Αν η πιθανότητα ενός ενδεχομένου είναι p η συμπληρωματική είναι $1 - p$.

ΟΙ ΛΟΓΙΣΤΙΚΕΣ ΜΟΝΑΔΕΣ (LOGIT)

Επειδή η πιθανότητα p είναι δύσκολο να μοντελοποιηθεί, αφού παίρνει τιμές στο κλειστό διάστημα μεταξύ του 0 και του 1, χρησιμοποιούμε συχνά στη στατιστική το κλάσμα της πιθανότητας προς την συμπληρωματική της. Για παράδειγμα αν ρίχνουμε ένα ζάρι με πιθανότητα να κερδίσουμε 50 τοις εκατό ή 0,5 τα odds είναι $\frac{0,5}{1-0,5} = \frac{0,5}{0,5} = 1$. Το odd λοιπόν είναι

$\frac{p}{1-p}$ και είναι η πιθανότητα ενός ενδεχομένου προς τη συμπληρωματική της. Τα κλάσμα αυτό

παίρνει τιμές από το 0 μέχρι θεωρητικά το συν άπειρο. Όσο η πιθανότητα ενός ενδεχομένου προσεγγίζει το μηδέν, το κλάσμα προσεγγίζει το μηδέν. Όσο η πιθανότητα προσεγγίζει τη μονάδα, το κλάσμα προσεγγίζει το συν άπειρο. Στο παράδειγμα που είδαμε στην προηγούμενη ενότητα, ο το κλάσμα της πιθανότητας προς τη συμπληρωματική της θα ήταν $\frac{\text{πιθανότητα ορθής απάντησης}}{\text{πιθανότητα λανθασμένης απάντησης}} = \frac{0,40}{0,60} \approx 0,66$. Για να μετατρέψουμε το κλάσμα πάλι σε

$$\text{πιθανότητα εργαζόμαστε αντίστροφα και έχουμε ότι } p = \frac{\text{odd}}{1 + \text{odd}} = \frac{0,66}{1 + 0,66} = 0,40$$

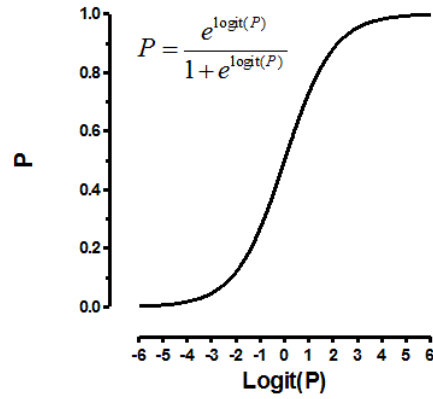
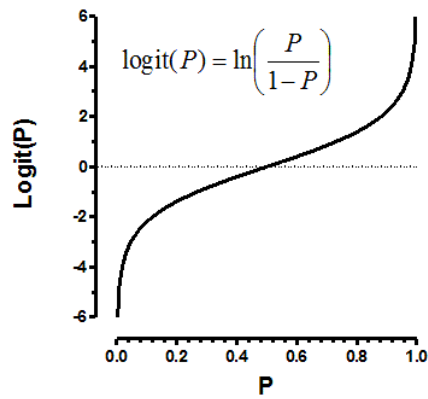
Το κλάσμα της πιθανότητας προς τη συμπληρωματική της έχει πάντα θετικές τιμές. Στον **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** που ακολουθεί βλέπουμε στην πρώτη γραμμή τις πιθανότητες από 0.01 ως 0,99. Στη δεύτερη γραμμή παρουσιάζονται τα odds δηλαδή οι λόγοι των πιθανοτήτων που βρίσκονται στη πρώτη γραμμή προς τις αντίστοιχες συμπληρωματικές πιθανότητες. Στη τρίτη γραμμή τιμές βλέπουμε ξανά τις τιμές odds σε δεκαδική μορφή να είναι μικρότερες της μονάδες για πιθανότητες μικρότερες του 0,5 και μεγαλύτερες από τη μονάδα για πιθανότητες μεγαλύτερες του 0,5. του κλάσματος αυτού για πιθανότητες από 0,01 ως 0,99 ή 99 τοις εκατό. Στην τέταρτη γραμμή παρουσιάζονται οι φυσικοί λογάριθμοι των odds δηλαδή ως δυνάμεις στις οποίες πρέπει να υψωθεί ο αριθμός e για να δώσει την τιμή του odd. Ο αριθμός e είναι ο 2,718. Για το παράδειγμα που αναφέραμε προηγουμένως έχουμε στην έβδομη στήλη του Πίνακας 1 ότι $e^{-0,41} = 0,66$.

Πίνακας 1. Odds (πιθανότητα προς τη συμπληρωματική της) και λογιστικές μονάδες.

Πιθανότητα	p	0,01	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
Odds	$\frac{p}{1-p}$	$\frac{1}{99}$	$\frac{10}{90}$	$\frac{20}{80}$	$\frac{30}{70}$	$\frac{40}{60}$	$\frac{50}{50}$	$\frac{60}{40}$	$\frac{70}{30}$	$\frac{80}{20}$	$\frac{90}{10}$
Τιμές των odds		0,01	0,11	0,25	0,43	0,66	1	1,50	2,33	4	9
logit	$\ln\left(\frac{p}{1-p}\right)$	-4,60	-2,20	-1,38	-0,84	-0,41	0	0,40	0,84	1,38	2,18

Με μαθηματικό συμβολισμό αυτό γράφεται ως $\ln\left(\frac{p}{1-p}\right) = y$ και $e^y = \left(\frac{p}{1-p}\right)$. Στο δικό μας παράδειγμα, έχουμε $0,66 = e^{-0,40}$ και συνεπώς $\ln(0,67) = -0,40$.

Ο μετασχηματισμός αυτός, όπως βλέπουμε στο Σχήμα 1, μετατρέπει την πιθανότητα στον κάθετο άξονα, η οποία είναι πάντα θετική και περιορίζεται στο διάστημα (0, 1), σε λογιστικές μονάδες που θεωρητικά παίρνουν τιμές από το πλην άπειρο στο συν άπειρο.



Στο κεφάλαιο αυτό λοιπόν θα διαβάσουμε την έρευνα των Σιδεριδη κ.ά. (2008), η οποία αφορά τη μελέτη μιας ορθογραφικής δοκιμασίας σε μαθητές Δημοτικού Σχολείου και η οποία δημοσιεύτηκε στο περιοδικό *Ψυχολογία* το 2008

Ψυχομετρική διερεύνηση μιας ορθογραφικής δοκιμασίας για μαθητές του δημοτικού σχολείου

ΓΕΩΡΓΙΟΣ Δ. ΣΙΔΕΡΙΔΗΣ¹, ΑΓΓΕΛΙΚΗ ΜΟΥΖΑΚΗ,
ΑΘΑΝΑΣΙΟΣ ΠΡΩΤΟΠΑΠΑΣ & ΠΑΝΑΓΙΩΤΗΣ ΣΙΜΟΣ

ΠΕΡΙΛΗΨΗ

Ο σκοπός αυτής της εργασίας ήταν η ψυχομετρική αξιολόγηση μιας νέας δοκιμασίας ορθογραφικής επίδοσης. Συμμετείχαν 580 μαθητές και μαθήτριες Β' - Δ' τάξης του Δημοτικού. Τα αποτελέσματα έδειξαν ότι υπήρχε μία μόνο εννοιολογική διάσταση της ορθογραφικής ικανότητας η οποία επιβεβαιώθηκε κατά τάξη και διαφορετικές χρονικές στιγμές. Οι δείκτες αξιοπιστίας και εσωτερικής συνοχής ήταν ικανοποιητικοί δείχνοντας τη σταθερότητα της κλίμακας. Ένας δεύτερος στόχος της παρούσας μελέτης ήταν η αξιολόγηση της ύπαρξης προκατάληψης της κλίμακας σε διαφορετικούς πολιτισμικά πληθυσμούς. Χρησιμοποιώντας αναλύσεις διαφορικής λειτουργίας των επιμέρους λέξεων (Differential Item Functioning) αξιολογήθηκε αν συγκεκριμένες λέξεις του τεστ ήταν σημαντικά πιο δύσκολες στους μαθητές που η μητρική τους γλώσσα δεν ήταν η ελληνική. Χρησιμοποιώντας το μοντέλο Rasch τα αποτελέσματα έδειξαν ότι η ύπαρξη προκατάληψης αφορούσε μόνο έναν πολύ μικρό αριθμό λέξεων του τεστ.

Λέξεις-κλειδιά: Ορθογραφία, Στάθμιση, Μέθοδος Rasch, Μεροληψία

1. Εισαγωγή

Η μελέτη της ορθογραφικής δεξιότητας παραμένει υποτιμημένη παρά τη σημαντική ακαδημαϊκή και γνωστική της διάσταση. Αν και η ψυχολογία της ανάγνωσης έχει καταγράψει θεμελιακά ευρήματα τις τελευταίες δεκαετίες, το επιστημονικό ενδιαφέρον επικεντρώνεται κυρίως στην αναγνωστική διαδικασία με συγκυριακές κυρίως αναφορές στην ορθογραφική δεξιότητα, η οποία συχνά αντιμετωπίζεται ως άμεσο παράγωγο της

εκμάθησης της ανάγνωσης. Παράλληλα, οι ερμηνείες των δεδομένων που έχουμε στη διάθεσή μας συγκλίνουν στη διαπίστωση ότι η ορθογραφική γραφή προϋποθέτει τη συνέργεια σημαντικών γνωστικών δεξιοτήτων οι οποίες αναπτύσσονται βραδύτερα από τις αναγνωστικές (Holmes & Babauia, 2005, Caravolas, Hulme & Snowling, 2001) και δεν χαρακτηρίζουν όλους ανεξαιρέτως τους καλούς αναγνώστες (Moats, 2005), υποδεικνύοντας μερική αποσύνδεση της αναγνωστικής διαδικασίας από την ορθογραφική

Οι Σιδερίδης κ.ά. (2008) πραγματοποίησαν έρευνα με σκοπό να μελετήσουν τα ψυχομετρικά χαρακτηριστικά μιας ορθογραφικής δοκιμασίας για μαθητές Δημοτικού. Στην πρώτη σελίδα της εργασίας των Σιδερίδη και των συνεργατών του (ό.π.) υπάρχει η περίληψη. Στην Εισαγωγή οι συγγραφείς αναφέρονται στη σημαντικότητα της μελέτης τους και εξηγούν ποιο κενό έρχονται να καλύψουν στη σχετική βιβλιογραφία.

1. Διεύθυνση επικοινωνίας: Γεώργιος Δ. Σιδερίδης, Τμήμα Ψυχολογίας, Πανεπιστήμιο Κρήτης, 74100 Ρέθυμνο Τηλ.: 28310-77545, e-mail:sideridis@psv.soc.uoc.gr

(Cossu, Gugliotta & Marshall, 1995). Όμως οι μετρήσεις αναγνωστικής και ορθογραφικής ακρίβειας φορτίζουν κατά μεγάλο μέρος σε κοινό παράγοντα-διάσταση αναγνωστικής δεξιότητας (και στα ελληνικά: Protorapas & Skaloumbakas, 2007. Πρωτόπαπας & Σκαλούμπakas, στο παρόν τεύχος) και οι μελετητές πιθανολογούν την ύπαρξη μιας κοινής πηγής για την ανάκληση λεξικών αναπαραστάσεων για την ανάγνωση και την ορθογραφική γραφή (Holmes & Babauta, 2005). Στον πυρήνα αυτής της πηγής βρίσκονται οι συγχωνευμένες (amalgamated) φωνολογικές και ορθογραφικές ταυτότητες των λέξεων, οι οποίες υποστηρίζουν τη διαδικασία της ανάγνωσης και της ορθογραφίας, ενώ ταυτόχρονα ισχυροποιούν τις σχετικές αναγνωστικές και ορθογραφικές δεξιότητες με την κάθε επαναλαμβανόμενη πρόσβαση και ανάλυση των εν λόγω μνημονικών αναπαραστάσεων (Ehri, 2005). Σύμφωνα με τη Snow και τους συνεργάτες της (2005), η ορθογραφία και η ανάγνωση στηρίζονται και εξαρτώνται από τις ίδιες νοητικές λεκτικές αναπαραστάσεις. Η γνώση της ορθογραφίας των λέξεων στο πλαίσιο ενός ορθογραφικού συστήματος υποδηλώνει στέρεες λεκτικές αναπαραστάσεις που εξυπηρετούν τη λεκτική αναγνώριση και προωθούν την αναγνωστική ευχέρεια (Snow et al., 2005). Κατά αυτό τον τρόπο, η ανάγνωση και η ορθογραφία θεωρούνται δύο αντίθετες όψεις μιας διαδικασίας που στηρίζεται σε βασικές γλωσσικές και μεταγλωσσικές δεξιότητες με κοινό πυρήνα τη φωνολογική επεξεργασία.

Με τον όρο *φωνολογική επεξεργασία* αναφερόμαστε στη σαφή κατανόηση και τον ενσυνείδητο χειρισμό των φωνολογικών μερών του λόγου σε επίπεδο συλλαβής και φωνήματος, πέρα από το νόημα και την επικοινωνιακή τους φύση. Το επίπεδο ανάπτυξης της φωνολογικής επίγνωσης κατά την είσοδο στο σχολείο έχει συσχετισθεί εδώ και αρκετό καιρό με την επιτυχία στη μάθηση της ανάγνωσης και της ορθογραφίας στις μεγαλύτερες τάξεις (Blachman, 1984. Bradley & Bryant, 1983. Juel, Griffith & Gough, 1986. Lundberg, Olofsson & Wall, 1980). Στην αγγλική γλώσσα, δεξιότητες φωνημικής κατάταξης προβλέπονται

σημαντική ακρίβεια την ορθογραφική ικανότητα στις πρώτες τάξεις του Δημοτικού σχολείου (Nation & Hulme, 1997). Στην ελληνική γλώσσα, η σχέση μεταξύ φωνολογικής επίγνωσης και ορθογραφίας φαίνεται να είναι αμοιβαία, όπως διαπιστώνεται από μελέτες στις οποίες τα παιδιά που κάνουν λιγότερα ορθογραφικά λάθη έχουν υψηλότερο επίπεδο φωνημικής επίγνωσης και, αντίστροφα, παιδιά που έχουν υψηλότερο επίπεδο φωνημικής επίγνωσης κάνουν λιγότερα ορθογραφικά λάθη (Πόρποδας, 1992. Porpodas, 1999. Αϊδίνης, 2006. Αϊδίνης & Παράσχου, 2004).

Το ελληνικό ορθογραφικό σύστημα

Το ελληνικό ορθογραφικό σύστημα θεωρείται σχετικά διαφανές (ρηχό), όσον ως προς την ανάγνωση, καταλαμβάνοντας τη δεύτερη θέση στην ταξινόμηση των Seymour και συνεργατών (2003), μετά το φινλανδικό, διότι υπάρχει συστηματική, συχνά ένα προς ένα, αντιστοιχία μεταξύ γραμμάτων και φθόγγων. Η συστηματικότητα της αντιστοιχίας αυτής χαρακτηρίζεται από υψηλή ασυμμετρία μεταξύ ανάγνωσης και ορθογραφίας. Ένας αρχάριος αναγνώστης μπορεί να διαβάσει με ακρίβεια την πλειονότητα των λέξεων, βασισμένος σε απλή εφαρμογή των γραφοφωνημικών αντιστοιχιών. Είναι όμως αδύνατο να τις γράψει σωστά εάν βασιστεί απλώς στην προφορά τους, διότι η ορθή γραφή των λέξεων της ελληνικής εξαρτάται σε μεγάλο βαθμό από τη μορφολογία και την ετυμολογία τους, η οποία συχνά βασίζεται στην ιστορική προέλευση της λέξης και μπορεί να είναι αρκετά σύνθετη.

Συγκεκριμένα, η μετατροπή των φθόγγων σε γραφήματα δεν είναι προβλέψιμη για τα 3 από τα 5 φωνήεντα, τα οποία μπορούν να γραφούν με ποικίλους τρόπους χρησιμοποιώντας ένα ή και δύο γράμματα. Έτσι, ενώ το /a/ γράφεται πάντοτε «α» και το /u/ πάντοτε «ου», υπάρχουν 6 γραφές για το /i/ (ι, η, υ, ει, οι, υι), 2 γραφές για το /o/ (ο και ω) και 2 για το /ε/ (ε και αι). Η σωστή ορθογραφία εξαρτάται από τη μορφολογία και την ετυμολογία της λέξης (Chliounaki & Bryant, 2002. Porpodas, 1999). Συνεπώς, για την

ΕΠΙΣΚΟΠΗΣΗ ΤΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

Στη δεύτερη σελίδα της εργασίας τους οι συγγραφείς προχωρούν σε εννοιολογικούς ορισμούς όπως είναι για παράδειγμα, η «ανάκληση λεξικών αναπαραστάσεων», οι «συγχωνευμένες φωνολογικές και ορθογραφικές ταυτότητες των λέξεων», το «επίπεδο φωνημικής επίγνωσης» και αναφέρονται σε άλλες σχετικές έρευνες στην βιβλιογραφία.

Στην επόμενη ενότητα οι συγγραφείς αναφέρονται στο ελληνικό ορθογραφικό σύστημα σε σχέση με την κατάταξη που έχει αυτό στην ελληνική ταξινόμηση του Seymour και των συνεργατών του και επισημαίνουν την «υψηλή ασυμμετρία» μεταξύ ανάγνωσης και ορθογραφίας που έχει η ελληνική γλώσσα.

ορθογραφημένη γραφή απαιτείται αναφορά σε γλωσσικά επίπεδα πέραν του φωνητικού. Λόγου χάρι, η λέξη «μήλο» γράφεται με τελικό «ο» επειδή πρόκειται για ουδέτερο ουσιαστικό, ενώ η λέξη «μιλώ» γράφεται με τελικό «ω» επειδή πρόκειται για ρήμα. Όμως οι ίδιες λέξεις διαφέρουν και στο θέμα («η» έναντι «ι»). Αυτή η διαφορά οφείλεται στην ετυμολογία τους και δεν μπορεί να προβλεφθεί ούτε φωνολογικά ούτε μορφολογικά. Απαιτεί γνώση των συγκεκριμένων μορφημάτων, γνώση που αναγκαστικά πρέπει να προσέλθει από την εξοικείωση με το γραπτό λόγο.

Η γραφημική απόδοση των συμφώνων είναι συνήθως προβλέψιμη αλλά όχι πάντοτε απλή. Τα φωνούμενα κλειστά (/b/, /d/, /g/, /j/) γράφονται με συνδυασμούς δύο γραμμάτων, συνδυάζοντας τα γράμματα για το ένρινο και το άφωνο κλειστό του ίδιου τόπου άρθρωσης. Έτσι το /b/ γράφεται μπ («καμπάνα»), όπως και το /mp/ («σαμπάνια»). Για το /j/ υπάρχει και εναλλακτική γραφή με «γγ». Αν και στην ελληνική φωνολογία δεν υπάρχουν μακρά ή διπλά σύμφωνα, ορισμένα σύμφωνα γράφονται διπλά λόγω ιστορικής ορθογραφίας (π.χ. το /m/ στο «λήμμα»). Το σύμφωνα /v/ και /f/ μπορούν να γραφούν, εκτός από τα αντίστοιχα γράμματα β και φ, με το γράμμα υ ή με συνδυασμό «υβ»/«υφ» σε ορισμένες λέξεις όταν προηγείται α ή ε (π.χ. «αύριο», «εύφορος»). Το /z/ γράφεται συνήθως με «ζ», αλλά όταν ακολουθεί φωνούμενο σύμφωνο γράφεται με «σ» (π.χ. «ασβός», «άσμα»). Τα συμφωνικά συμπλέγματα /ks/ και /ps/ γράφονται με τα γράμματα «ξ» και «ψ», αντίστοιχα. Τα ουρανικά κλειστά (/c/, /j/) και τριβόμενα (/ç/, /j/) σύμφωνα συνήθως γράφονται με το γράμμα του όμοιου σε τρόπο και φώνηση υπερωικού φθόγγου, ακολουθούμενο από ένα από τα 6 γραφήματα που αντιστοιχούν στο /i/ (π.χ. «κιούπι», «φτώχεια»), εκτός αν ο επόμενος φθόγγος είναι /e/, οπότε το /i/ παραλείπεται (π.χ. «κέφι», «χέρι»). Αν όμως προηγείται σύμφωνο, ένα ουρανικό τριβόμενο μπορεί να γραφτεί και μόνο με γράφημα /i/ χωρίς συμφωνικό γράμμα (π.χ. «κάποια», «βιάσου»). Το ουρανικό πλάγιο γράφεται με «λ» ακολουθούμενο από γράφημα του /i/ (π.χ. «λιάζω», «μακλειό»), ενώ το ουρα-

νικό ένρινο γράφεται με «ν» ακολουθούμενο από γράφημα του /i/ (π.χ. «πανιά», «νοιάζει»), εκτός αν προηγείται /m/ (π.χ. «ζημιά», «μοιάζω»). Τέλος, υπάρχουν λίγες περιπτώσεις γραμμάτων χωρίς αντίστοιχο φθόγγο (το «π» σε συμπλέγματα μεταξύ «μ» και «τ», όπως «πέμπτος»).

Συνοψίζοντας, το ελληνικό ορθογραφικό σύστημα χαρακτηρίζεται από πολυσήμαντη αντιστοιχία (όπου ένας φθόγγος αντιστοιχεί σε περισσότερα από ένα γράφηματα), ταυτόχρονη αντιστοιχία δύο ή περισσότερων φθόγγων στο ίδιο γράφημα, και σε μικρό βαθμό αναντιστοιχία (όπου γίνεται χρήση γραφημάτων χωρίς αντίστοιχο φθόγγο) (Porrodas, 2001. Καρατζάς, 2005). Συνεπώς, η ελληνική ορθογραφία είναι σύνθετη και παρουσιάζει ιδιαιτερότητες, ιδιαίτέρως εκτεταμένες στη χρήση φωνηέντων, εξαιτίας των διάφορων μορφολογικών κατηγοριών και της ιστορικής εξέλιξης των λέξεων. Γι' αυτό, η ακριβής γραφή των λέξεων εξαρτάται, εκτός από τη φωνημική επίγνωση του αναγνώστη, και από το οπτικοορθογραφικό του λεξικό και τη μορφολογική του επίγνωση.

Εκτίμηση της ορθογραφικής δεξιότητας

Μεγάλο μέρος της ανάπτυξης των ορθογραφικών δεξιοτήτων παρατηρείται κατά τη διάρκεια της βασικής εκπαίδευσης όπου η διδασκαλία ορθογραφικών συμβάσεων, ετυμολογίας και γραμματικών κανόνων αποσκοπεί αφενός στην απόκτηση γνώσεων γύρω από το ορθογραφικό σύστημα και τα χαρακτηριστικά του και αφετέρου στην ανάπτυξη μεταγλωσσικών δεξιοτήτων. Τέτοιες δεξιότητες, που σχετίζονται κυρίως με το συλλογισμό σχετικά με τις συμβάσεις του γραπτού λόγου, εμπλέκονται αποφασιστικά στη διαδικασία της παραγωγής του και επηρεάζουν τις επιδόσεις των μαθητών σε αντίστοιχα γνωστικά έργα. Σε περιορισμένο αριθμό μελετών που έχουν διεξαχθεί με μαθητές του Δημοτικού σχολείου έχει παρατηρηθεί πρώιμη χρήση φωνολογικών στρατηγικών τόσο σε φωνημικό όσο και σε συλλαβικό επίπεδο (Porrodas, 2001). Οι συγκεκριμένες στρατηγικές συμπληρώνονται βαθμιαία από μορφολογικές στρατηγικές και με αυτό τον

Οι συγγραφείς συνοψίζουν στη σελίδα 292 της εργασίας τους ότι το ελληνικό ορθογραφικό σύστημα χαρακτηρίζεται από πολυσήμαντη αντιστοιχία, ταυτόχρονη αντιστοιχία αλλά και αναντιστοιχία μεταξύ φθόγγων και γραφημάτων. Έτσι καταλήγουν στο συμπέρασμα ότι «η ακριβής γραφή των λέξεων εξαρτάται, εκτός από τη φωνημική επίγνωση του αναγνώστη, και από το οπτικοορθογραφικό του λεξικό και τη μορφολογική του επίγνωση».

Οι συγγραφείς παρουσιάζουν στοιχεία για παρόμοιες διδακτικές παρεμβάσεις στην Ελλάδα και στην δεξιά στήλη της επόμενης σελίδας 293 αναφέρονται στη δική τους εργασία ως μέρος μιας ευρύτερης διαχρονικής μελέτης που περιελάμβανε πέντε διαδοχικές μετρήσεις.

τρόπο επιτρέπουν την ορθότερη γραφή γραμματικών μορφημάτων (Nunes, Aidinis & Bryant, 2006), καθώς φαίνεται η συχνότητα εμφάνισης μορφημάτων να συμβαδίζει με αποτελεσματικότερη εφαρμογή μορφολογικών στρατηγικών (Διακογιώργη, Μπαρής & Βαλμάς, 2005). Επιπλέον, μια πρόσφατη μελέτη από τους Νικολοπούλου, Goulandris & Snowling (2003) σε δείγμα 28 δυσλεξικών παιδιών κατέδειξε σημαντικές διαφορές στις ορθογραφικές δεξιότητες μεταξύ των Ελλήνων δυσλεξικών μαθητών και της ομάδας ελέγχου που αποτελούσαν συνομήλικοί τους μαθητές μέσης αναγνωστικής επίδοσης, ενώ οι διαφορές μεταξύ των δύο ομάδων στην αναγνωστική ακρίβεια δεν ήταν σημαντικές. Η πλειονότητα των λαθών ορθογραφίας των παιδιών με δυσλεξία ήταν συνήθως μορφολογικής φύσης, ενώ δεν υπήρξε καμία περίπτωση φωνολογικού λάθους. Τέλος, τα δυσλεξικά παιδιά βρέθηκαν να έχουν ιδιαίτερες δυσκολίες στην ορθογραφία κλιτικών μορφημάτων με πολλά γράμματα, εύρημα που αποκαλύπτει ιδιαίτερη αδυναμία στη γνώση ορθογραφικών συνδυασμών.

Αντίστοιχες διδακτικές παρεμβάσεις που στοχεύουν στην ανάπτυξη ορθογραφικών δεξιοτήτων συνήθως βασίζονται στην καλλιέργεια μεταγλωσσικών δεξιοτήτων σε φωνολογικό και μορφολογικό επίπεδο, συχνά όμως δεν συνοδεύονται από κριτήρια προσδιορισμού της αποτελεσματικότητάς τους. Ο προσδιορισμός του βαθμού ανάπτυξης της ορθογραφικής δεξιότητας μέσω ποσοτικών κριτηρίων (νόρμες) ή μέσω ποιοτικών διαδικασιών (ανάλυση της ορθής και της λανθασμένης γραφής επιμέρους λέξεων) μπορεί να αποβεί εξαιρετικά χρήσιμος γιατί μας βοηθά να εκτιμήσουμε τη γενικότερη εξοικείωση του μαθητή με το γραπτό λόγο και την ειδικότερη γνώση του για συγκεκριμένες πτυχές του ορθογραφικού συστήματος. Στον ελλαδικό χώρο όμως διαπιστώνεται έλλειψη κατάλληλων εργαλείων αξιολόγησης της ορθογραφίας, με αποτέλεσμα οι κύριες προσεγγίσεις για την εκτίμηση του επιπέδου ανάπτυξης της ορθογραφικής δεξιότητας (τόσο σε τυπικό πληθυσμό όσο και στους μαθητές με μαθησιακές δυσκο-

λίες) να βασίζονται κυρίως στην προσωπική εκπαιδευτική ή κλινική εμπειρία του κάθε εξεταστή/ερευνητή.

Η υπό διερεύνηση ορθογραφική δοκιμασία αποτελεί μέρος ενός πρόσφατου εγχειρήματος συγκέντρωσης πληροφοριών για την ανάπτυξη της αναγνωστικής και ορθογραφικής δεξιότητας από ένα επαρκές δείγμα ελληνικού μαθητικού πληθυσμού του δημοτικού σχολείου. Η διαχρονική αυτή μελέτη διήρκεσε τρία σχολικά έτη (2004-2007) και περιλάμβανε 5 διαδοχικές επιτρήσεις σε τακτά χρονικά διαστήματα (περίπου 6 μηνών). Έτσι, το αρχικό δείγμα των μαθητών της μελέτης επαναξιολογήθηκε στην πλειονότητά του (89,5%) για τέσσερις ακόμη φορές επιτρέποντας: (α) την αξιολόγηση της σταθερότητας (αξιοπιστίας επανεξέτασης) των μετρήσεων, και (β) τον προσδιορισμό της προβλεπτικής ισχύος της επίδοσης στη συγκεκριμένη δοκιμασία ορθογραφίας σε σχέση με την επίδοση των ίδιων μαθητών σε μια σειρά αναγνωστικών δοκιμασιών (τρέχουσα εγκυρότητα). Τα ευρήματα από τη συγκεκριμένη διαχρονική μελέτη έχουν συνεισφέρει στην πληρέστερη κατανόηση της ανάπτυξης δεξιοτήτων ανάγνωσης και ορθογραφίας στο ελληνικό δημοτικό σχολείο (Mouzaki & Sideridis, 2007. Protopapas et al., 2007. Sideridis et al., 2006) προσφέροντας ταυτόχρονα δεδομένα στάθμισης για τα ψυχομετρικά εργαλεία που χρησιμοποιήθηκαν. Μέρος των δεδομένων που αναφέρονται στην εν λόγω δοκιμασία για την εκτίμηση της ορθογραφικής δεξιότητας προερχόμενα από τις τρεις πρώτες μετρήσεις και τις μεσαίες τάξεις του δημοτικού έχουν αναλυθεί προηγουμένως με ικανοποιητικά αποτελέσματα ως προς τα ψυχομετρικά της χαρακτηριστικά (Μουζάκη κ.ά., 2007). Στην παρούσα μελέτη περιλαμβάνονται νέα δεδομένα που προέρχονται από την πέμπτη επαναληπτική μέτρηση του αρχικού δείγματος των μαθητών και από τα οποία συνάγονται οι μέσες και εκατοστιαίες τιμές (νόρμες) για την τελευταία (Στ') τάξη του Δημοτικού σχολείου. Επιπλέον, εξετάζονται εκ νέου τα ψυχομετρικά χαρακτηριστικά της δοκιμασίας με τη χρήση πιο σύνθετων ανα-

βιβλιογραφική επισκόπηση συνεχίζει με τις έννοιες της «αξιοπιστίας επανεξέτασης» και της «τρέχουσας εγκυρότητας». Η εγκυρότητα ενός ερευνητικού εργαλείου είναι ο βαθμός που αυτό αντανακλά την προς μέτρηση θεωρητική κατασκευή (εδώ η «ορθογραφική δεξιότητα»). Η εγκυρότητα είναι «τρέχουσα» (αλλιώς «σύγχρονη») όταν αυτή αναφέρεται στην στενή σχέση ενός τεστ ορθογραφικής δεξιότητας με ένα άλλο τεστ ορθογραφικής δεξιότητας για το οποίο γνωρίζουμε ότι μετράει αυτή ακριβώς τη θεωρητική κατασκευή. Η αξιοπιστία, από την άλλη μεριά, αναφέρεται στον βαθμό που ένα τεστ δίνει ίδια αποτελέσματα για το ίδιο ποσό της θεωρητικής κατασκευής. Για παράδειγμα, ένα θερμόμετρο είναι αξιόπιστο αν δίνει την ίδια μέτρηση για την ίδια

θερμοκρασία. Η αξιοπιστία συνοδεύεται με τον όρο «επανεξέτασης» όταν ελέγχεται με την επίδοση ενός παρόμοιου τεστ. Για παράδειγμα ένα τεστ είναι αξιόπιστο όταν με αυτό επανεξετάζεις τους ίδιους μαθητές και παίρνεις τα ίδια αποτελέσματα. Δηλαδή όσοι βαθμολογήθηκαν με υψηλό βαθμό στο ένα τεστ έχουν βαθμολογηθεί με υψηλό βαθμό και στο άλλο κα -αντιστρόφως- όσοι έχουν βαθμολογηθεί με χαμηλό βαθμό στο ένα έχουν βαθμολογηθεί με χαμηλό βαθμό και στο άλλο.

Πίνακας 1
Κατανομή του δείγματος κατά τάξη και φύλο και
ηλικία σε μήνες (ΜΟ±ΤΑ και εύρος σε παρένθεση)

	Τάξη	Β'	Γ'	Δ'	Ε'	Στ'
Αγόρια	N	95	85	85	75	73
	Ηλικία	93±3,5 (87-99)	105±3,3 (99-111)	117±3,6 (111-123)	129±3,6 (123-135)	141±3,6 (135-147)
Κορίτσια	N	101	96	91	81	79
	Ηλικία	93±3,2 (87-99)	105±3,4 (96-111)	117±3,4 (111-123)	129±3,4 (123-135)	141±3,4 (135-147)
Συνολικό	N	196	181	176	156	142

λυτικών διαδικασιών για τον προσδιορισμό αφενός της κατασκευαστικής/δομικής εγκυρότητας και αφετέρου της ύπαρξης πιθανής μεροληψίας. Το ερευνητικό μας ενδιαφέρον για το συγκεκριμένο θέμα υπαγορεύεται τόσο από τη σπουδαιότητα που αποδίδεται στην ορθογραφία ως γνωστική διαδικασία στο χώρο του σχολείου όσο και από την έλλειψη εμπειρικών δεδομένων που να διευκρινίζουν το εύρος των ορθογραφικών δεξιοτήτων των μαθητών στο ελληνικό Δημοτικό σχολείο.

2. Μέθοδος

Συμμετέχοντες

Η δοκιμασία της ορθογραφίας χορηγήθηκε ατομικά σε 587 μαθητές 17 Δημοτικών σχολείων από 3 γεωγραφικές περιφέρειες της χώρας (Κρήτη, Αττική και Επύρνησα). Τα σχολεία επιλέχθηκαν με μη συστηματικό τρόπο, ο οποίος εξυπηρέτούσε τη συλλογή δεδομένων από τις συγκεκριμένες περιοχές και την αντιπροσώπευση αστικών, ημιαστικών και αγροτικών σχολικών μονάδων. Συγκεκριμένα, 7 σχολεία ήταν αστικά (41%), 6 ημιαστικά (35%) και 4 αγροτικά (24%). Δεδομένα για τις τάξεις Β', Γ' και Δ' συνελέγησαν το Μάιο του 2005 από μαθητές οι οποίοι επιλέχθηκαν τυχαία από τα συγκεκριμένα σχολεία, εφόσον πληρούσαν τα εξής κριτή-

ρια: (α) δεν είχαν επαναλάβει κάποια από τις προηγούμενες τάξεις, (β) απουσία ιστορικού νευρολογικής ή ψυχιατρικής νόσου ή αισθητηριακής διαταραχής (σύμφωνα με τον εκπαιδευτικό της τάξης), και (γ) έγγραφη συναίνεση των γονέων. Οι μαθητές, δεδομένα από τους οποίους παρουσιάζονται εδώ, πληρούσαν ένα επιπλέον κριτήριο σχετικό με τη νοημοσύνη (βλ. παρακάτω). Δεδομένα για την Ε' και Στ' τάξη συνελέγησαν το Μάιο του 2006 και το Μάιο του 2007 κατά την επανεξέταση 156 και 152 μαθητών, αντίστοιχα, οι οποίοι κατά την πρώτη μέτρηση φοιτούσαν στην Δ' τάξη.

Η κατανομή του τελικού δείγματος κατά τάξη και φύλο παρουσιάζεται στον Πίνακα 1. Κατά την πρώτη μέτρηση 48 μαθητές προέρχονταν από οικογένειες αλλοδαπών (κυρίως από την Αλβανία, 20 στη Β', 15 στην Γ' και 13 στην Δ' Τάξη) οι οποίοι μιλούσαν ελληνικά σε ικανοποιητικό επίπεδο, σύμφωνα με τον εκπαιδευτικό της τάξης, και είχαν φοιτήσει στο ελληνικό σχολείο από την Α' Δημοτικού.

Μετρήσεις

Δοκιμασία ορθογραφίας. Η ορθογραφική δεξιοσύνη των μαθητών αξιολογήθηκε βάσει της γραφής μιας σειράς λέξεων επιλεγμένων από το βασικό λεξιλόγιο του αναγνωστικού μαθήματος για όλες τις τάξεις του Δημοτικού σχολείου. Συ-

Τέλος, όταν αναφερόμαστε σε «στάθμιση» εννοούμε τη διαδικασία στην οποία ένα τεστ (καινούργιο ή προσαρμοσμένο από άλλη γλώσσα) εξετάζεται σε άλλον πληθυσμό προκειμένου να δούμε αν το τεστ αυτό λαμβάνει υπόψη τα χαρακτηριστικά του πληθυσμού στον οποίο δίνεται. Στη σελίδα 294 οι συγγραφείς δίνουν στον Πίνακα 1 στοιχεία περιγραφικής στατιστικής για τους συμμετέχοντες στην έρευνα σε ένα πίνακα διπλής εισόδου κατά σχολική τάξη και φύλο. Η ηλικία των παιδιών εκφράζεται σε μήνες και όχι σε έτη, πράγμα που είναι αναγκαίο από ουσιαστική και ερευνητική άποψη. Σε κάθε κελί του Πίνακα 1 στην εργασία των Σιδερίδη κ. ά. (2008: 294) βλέπουμε το μέγεθος του δείγματος,

τον μέσο όρο ακολουθούμενο από την τυπική απόκλιση και το εύρος της κατανομής της ηλικίας.

ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Στη ενότητα «Μέθοδος» [συλλογής δεδομένων] και στην υποενότητα «Συμμετέχοντες» διαβάζουμε ότι τα σχολεία επιλέγηκαν με «μη συστηματικό τρόπο, ο οποίος εξυπηρετούσε τη συλλογή δεδομένων από τις συγκεκριμένες περιοχές ...». Αργότερα δίδονται τα ποσοστά των σχολείων σε αστικές, ημιαστικές και αγροτικές περιοχές, καθώς και τα κριτήρια που θα έπρεπε να πληρούν οι μαθητές των σχολείων αυτών για να συμμετάσχουν στην έρευνα. Ένα σημαντικό στοιχείο εδώ είναι η «έγγραφη συναίνεση των γονέων». Οι ερευνητές συμπεριέλαβαν στο δείγμα τους και 48 μαθητές που δεν είχαν την ελληνική ως μητρική γλώσσα, πράγμα που, όπως θα δούμε παρακάτω, είχε μεγάλη θεωρητική και στατιστική σημασία.

μπερλήφθηκαν ουσιαστικά, ρήματα, επίθετα, σύνδεσμοι, επιρρήματα, προθέσεις και μετοχές. Τα κριτήρια για την επιλογή των λέξεων (αντιπροσωπευση συγκεκριμένων μορφολογικών και γραμματικών χαρακτηριστικών) διαμορφώθηκαν σύμφωνα με τη σειρά με την οποία διδάσκονται τα διάφορα γραμματικά φαινόμενα στο αναλυτικό πρόγραμμα. Σωστή απόδοση της συμβατικής ορθογραφίας κάθε λέξης απαιτεί την εφαρμογή ποικίλων στρατηγικών (φωνολογικών, μορφολογικών, ετυμολογικών). Επιπλέον, η επιλογή κλιτών λέξεων που υπόκεινται σε καταληκτικούς κανόνες ή άλλες μορφολογικές συμβάσεις, και λέξεων των οποίων το θεματικό μέρος συνδέεται με την ετυμολογία της λέξης, επιτρέπουν την παραγωγή λαθών με σημαντική διαγνωστική αξία. Τα πιθανά ορθογραφικά λάθη επιτρέπουν τόσο ποσοτική όσο και ποιοτική ανάλυση, με στόχο να καταρτιστεί ένα περίγραμμα δυνατοτήτων και αδυναμιών του μαθητή βάσει του οποίου θα σχεδιαστεί η κατάλληλη διδακτική παρέμβαση. Οι λέξεις ιεραρχήθηκαν ανάλογα με τον αριθμό συλλαβών, ορθογραφικών χαρακτηριστικών/μορφημάτων και το βαθμό ορθογραφικής δυσκολίας τους σύμφωνα με τις κρίσεις ομάδας εμπειρων εκπαιδευτικών στη φάση της κατασκευής του εργαλείου.

Στο πλαίσιο της δοκιμασίας η κάθε λέξη υπαγορεύεται στο μαθητή, συνοδευόμενη από μια σύντομη πρόταση που περιέχει τη συγκεκριμένη λέξη. Για τη γραφή των λέξεων μπορεί να χρησιμοποιηθεί οποιοδήποτε φύλλο γραφής με αριθμημένες σειρές. Ο εξεταστής παρακολουθεί τη διαδικασία και σταματά τη χορήγηση όταν ο μαθητής έχει κάνει ορθογραφικά λάθη σε έξι διαδοχικές λέξεις (αγνοώντας τυχόν λάθη στον τονισμό). Η μέγιστη δυνατή βαθμολογία είναι 60 (μία μονάδα για κάθε ορθογραφημένη λέξη).

Λεξιλόγιο και σχέδια με κύβους του WISC-III. Οι δοκιμασίες αυτές περιλαμβάνονται στην ελληνική έκδοση του Wechsler Intelligence Scale for Children III (Γεώργας κ.ά., 1997). Χορηγήθηκαν στο σύνολο του δείγματος με σκοπό την αδρή εκτίμηση του νοητικού επιπέδου των μαθητών.

3. Αποτελέσματα

Το τελικό δείγμα που παρουσιάζεται στον Πίνακα 1 περιλαμβάνει μαθητές με εκτιμώμενο δείκτη νοημοσύνης >70 , ο οποίος αντιστοιχεί σε μέσο όρο τυπικών τιμών στις κλίμακες Σχέδια με Κύβους και Λεξιλόγιο <4 .

Όπως έχει αναφερθεί και προηγουμένως (Μουζάκη κ.ά., 2007), ο βαθμός εσωτερικής αξιοπιστίας (συνέπειας) των επιμέρους λέξεων της δοκιμασίας της ορθογραφίας υπήρξε ικανοποιητικός (α του Cronbach=0,945, $N=548$). Αντίστοιχα πολύ υψηλή ήταν και η εσωτερική συνοχή της κλίμακας για κάθε τάξη (Β': $a=0,89$, Γ': $a=0,93$, Δ': $a=0,94$, Ε': $a=0,95$, ΣΤ': $a=0,94$). Υποθέτουμε ότι η αύξηση της τιμής του a ανάλογα με την ηλικία (τάξη) οφείλεται εν μέρει στη σταθεροποίηση των ορθογραφικών τους ικανοτήτων. Επίσης, η αύξηση αυτή μπορεί να αντανάκλα εν μέρει και το γεγονός ότι στον υπολογισμό του a στις μεγαλύτερες τάξεις λαμβάνονται υπόψη περισσότερες λέξεις οι οποίες δεν είχαν χορηγηθεί στις μικρότερες τάξεις λόγω του βαθμού δυσκολίας τους (κανόνες οροφής), γεγονός που είχε ως αποτέλεσμα μικρότερη διακύμανση.

Η συμβολή των επιμέρους λέξεων στη συνολική βαθμολογία της κλίμακας (σύμφωνα με το δείκτη συνάφειας λέξης-συνόλου) κρίνεται ικανοποιητική με τιμές $>0,50$ για τη συντριπτική πλειονότητα των λέξεων. Ένας μικρός αριθμός λέξεων παρουσιάζει μικρότερη συνάφεια (item-total μεταξύ 0,08 και 0,20) με τη συνολική βαθμολογία (οι τελευταίες 5-6 λέξεις για τους μαθητές της Β' τάξης και οι πρώτες 6-7 λέξεις για τους μαθητές των μεγαλύτερων τάξεων). Η συμπεριλήψή τους όμως στη δοκιμασία κρίνεται απαραίτητη (ιδιαίτερα των αρχικών) με σκοπό τη σταδιακή αντιμετώπιση λέξεων με σύνθετη ορθογραφία, γεγονός απαραίτητο κυρίως σε παιδιά μικρότερων ηλικιών ή παιδιών με δυσκολίες στο γραπτό λόγο.

Η σειρά κατάταξης των λέξεων βάσει δυσκολίας (πιθανότητα σωστής γραφής της κάθε λέξης στο σύνολο του δείγματος) βρίσκεται σε υψηλή συνάφεια με τη σειρά χορήγησης (ρ

σελίδα 295 οι Σιδερίδης κ.ά. (ο.π.) αναφέρουν τον τρόπο βαθμολόγησης της ορθογραφικής ικανότητας των μαθητών και τη διαδικασία του τεστ. Αναφέρουν επίσης το τεστ νοητικού επιπέδου του Wechsler. Αναφέρονται κατόπιν στον βαθμό εσωτερικής αξιοπιστίας (συνέπειας) των επιμέρους λέξεων της δοκιμασίας» και αναφέρουν ότι αυτός υπήρξε ικανοποιητικός (α του Cronbach).

Ο ΔΕΙΚΤΗΣ ΑΛΦΑ

Ο εν λόγω δείκτης προτάθηκε το 1951 από τον Αμερικανό ψυχολόγο και στατιστικό Lee Joseph Cronbach, ο οποίος, πέρα από τον δείκτη αυτόν, εισήγαγε και τη Θεωρία της Γενικευσιμότητας. Η κεντρική ιδέα του δείκτη άλφα είναι ότι αν τα επιμέρους ερωτήματα ενός τεστ μετρούν την ίδια θεωρητική κατασκευή τότε

αυτή θα «προκαλεί» υψηλό βαθμό συνάφειας ανάμεσα στα ερωτήματα. Από μαθηματική άποψη ο α είναι ένα κλάσμα που έχει στον αριθμητή το άθροισμα των διακυμάνσεων των επί μέρους ερωτημάτων και στον παρονομαστή τη συνολική διακύμανση (η ποσότητα του αριθμητή συν την συνδιακύμανση των ερωτημάτων). Το εν λόγω κλάσμα αφαιρείται από τη μονάδα και πολλαπλασιάζεται επί έναν παράγοντα που έχει να κάνει με το πλήθος των ερωτημάτων. Αν δεν υπάρχει συνδιακύμανση, το κλάσμα γίνεται μονάδα (ο αριθμητής είναι ίσος με τον παρονομαστή) και ο α γίνεται μηδέν αφού το κλάσμα αφαιρείται από τον αριθμό 1. Αν η συνδιακύμανση μεταξύ των ερωτημάτων είναι μεγάλη, ο παρονομαστής μεγαλώνει, το κλάσμα μικραίνει και ο δείκτης πλησιάζει τη μονάδα, αφού η ποσότητα που αφαιρείται από το 1 είναι μικρή. Αν έχουμε ανάμεικτη θετική και αρνητική συνδιακύμανση ανάμεσα στα ερωτήματα και παράλληλα αυτή η συνδιακύμανση είναι σχετικά μεγάλη (π.χ. άσχετα μεταξύ τους ερωτήματα ή κακές επιδόσεις), ο α μπορεί να εμφανιστεί και με αρνητικό πρόσημο. Πάντως, υψηλή και θετική τιμή του α , δηλαδή υψηλός βαθμός εσωτερικής αξιοπιστίας, δεν σημαίνει αναγκαστικά ότι έχουμε μία και μόνο θεωρητική κατασκευή στο τεστ. Τέλος ο δείκτης α έχει την ίδια λογική με εκείνη των δεικτών μεγέθους της επίδρασης γιατί αν αφαιρέσουμε από τη μονάδα το τετράγωνο του α παίρνουμε τη διακύμανση που δεν εξηγείται. Οι συγγραφείς στο κάτω δεξιό μέρος της σελίδας 295 παρουσιάζουν τους δείκτες συνάφειας μεμονωμένων ερωτημάτων (εδώ λέξεων) με το σύνολο του τεστ, ένα αποτέλεσμα που εμφανίζεται στο SPSS στη διαδικασία Reliability Analysis, και

Πίνακας 2
Μέσοι όροι (ΜΟ), τυπικές αποκλίσεις (ΤΑ), και εκατοστιαίες τιμές
για τον αριθμό σωστών απαντήσεων στη δοκιμασία της ορθογραφίας κατά τάξη και φύλο

Τάξη	Φύλο	N	ΜΟ (ΤΑ)	Εκατοστημόριο						
				5	10	25	50	75	90	95
Β'	Αγόρια	95	22,4 (8,0)	10	12	16	22	27	35	38
	Κορίτσια	101	23,8 (6,3)	12	16	20	24	28	32	35
Γ'	Αγόρια	85	31,0 (9,5)	14	18	24	31	38	45	48
	Κορίτσια	96	33,0 (8,5)	19	22	27	34	39	44	47
Δ'	Αγόρια	85	36,2 (10,0)	17	22	29	37	45	51	53
	Κορίτσια	91	39,6 (10,0)	22	27	33	41	47	51	53
Ε'	Αγόρια	75	40,1 (11,0)	18	25	33	43	49	53	55
	Κορίτσια	81	46,0 (10,0)	27	30	42	48	53	56	57
ΣΤ'	Αγόρια	73	43,0 (10,3)	23	29	36	46	52	54	56
	Κορίτσια	79	48,6 (8,8)	30	36	44	51	55	57	58

του Spearman=0,940). Με λίγες μεμονωμένες εξαιρέσεις, όπως οι λέξεις «πηγή» (αρ. 16), «φιλί» (αρ. 17) και «δίχτυ» (αρ. 19), η σειρά χορήγησης αντανάκλα με ακρίβεια την παρατηρούμενη δυσκολία στην ορθογραφία των λέξεων. Πολύ υψηλή ήταν η συνάφεια της σειράς κατάταξης μεταξύ των τριών τάξεων (Β' - Γ': $\rho=0,974$; Β' - Δ': $\rho=0,978$; Γ' - Δ': $\rho=0,976$, πρώτη μέτρηση), επιβεβαιώνοντας τη σταθερότητα της διαφοράς δυσκολίας μεταξύ των λέξεων και συνεπώς της διαφορικής τους χρησιμότητας στην εκτίμηση της ορθογραφικής δεξιότητας.

Ο βαθμός αξιοπιστίας επανεξέτασης της δοκιμασίας ήταν επίσης ικανοποιητικός αφού ο δείκτης συνάφειας μεταξύ της πρώτης και της δεύτερης μέτρησης της ορθογραφικής δεξιότητας που διενεργήθηκε ένα έτος αργότερα ήταν πολύ υψηλός (r του Pearson=0,91 με βάση αρχικές τιμές στο σύνολο του δείγματος).

Ο Πίνακας 2 περιέχει περιγραφικά στοιχεία για την κατανομή των ατομικών βαθμολογιών στη δοκιμασία ορθογραφίας για τις τάξεις Α' - ΣΤ'. Βρέθηκαν σημαντικές διαφορές μεταξύ των δύο φύλων (κύρια επίδραση της μεταβλητής Φύλο, $F[1, 550]=10,10, p<0,002$ στην πρώτη μέτρηση,

$F[1, 491]=14,77, p<0,0001$ στη δεύτερη μέτρηση και $F[1, 468]=20,98, p<0,0001$ στην τρίτη μέτρηση, όχι όμως και σημαντική αλληλεπίδραση Φύλου και Τάξης ($p>0,4$). Επομένως, τα τυπικά δεδομένα παρουσιάζονται ξεχωριστά για κάθε φύλο.

Ανίχνευση απλών δομών ή κατασκευαστικής/δομικής εγκυρότητας με ανάλυση κυρίων συνιστωσών (Principal Components Analysis)

Σε δεύτερη φάση, διερευνήθηκε η ύπαρξη μίας ή περισσότερων διαστάσεων ορθογραφικής ικανότητας. Η τυπική μέθοδος ανάλυσης είναι η ανάλυση παραγόντων με θεωρητικά πιο σωστή της μορφή την ανάλυση κυρίων συνιστωσών, η οποία είναι ικανή να «διακρίνει απλές δομές» με μεγαλύτερη σαφήνεια (Gorsuch, 1983). Θεωρήσαμε ότι οι 60 λέξεις θα πρέπει να περιγράφονται ικανοποιητικά από μία συνιστώσα διάσταση (μία εννοιολογική μεταβλητή). Εξαιτίας της παρουσίας διχοτομικών απαντήσεων (σωστό-λάθος), οι αναλύσεις μπορούν να πραγματοποιηθούν με δύο τρόπους: (α) με τον υπολογισμό πολυχωρικών συντελεστών (π.χ., με τη χρήση του

αναφέρουν ότι αυτοί είναι για όλες τις λέξεις πάνω από 0,50 (πράγμα φυσικό κατά τη γνώμη μας).

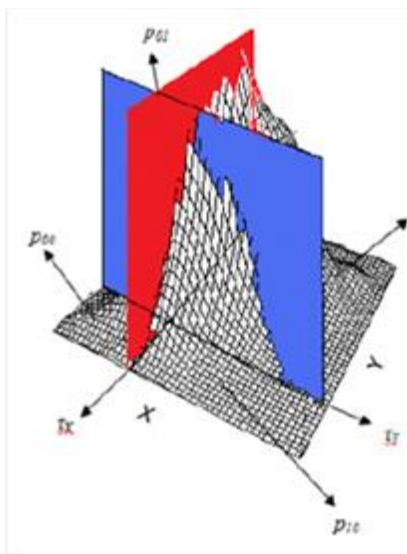
Ο ΔΕΙΚΤΗΣ r ΤΟΥ PEARSON ΚΑΙ Ο ΔΕΙΚΤΗΣ ρ ΤΟΥ SPEARMAN

Στη συνέχεια, οι ερευνητές χρησιμοποίησαν τον δείκτη ρ (ρ , rho) του Spearman για να διερευνήσουν αν η σειρά με την οποία επιδόθηκαν οι λέξεις στους μαθητές

ήταν σε συμφωνία με τον βαθμό δυσκολίας τους. Ο δείκτης ρ χρησιμοποιείται εναλλακτικά αντί για τον δείκτη r του Pearson όταν τα δεδομένα δεν έχουν κανονική κατανομή ή όταν η σχέση που τα συνδέει δεν είναι γραμμική ή όταν έχουμε τακτικές τιμές. Μάλιστα, ο Charles Edward Spearman, ο Άγγλος ψυχολόγος που πρότεινε τον εν λόγω δείκτη, έγινε γνωστός για τη θεωρία του περί ύπαρξης ενός γενικού παράγοντα νοημοσύνης, καθώς και για τις εργασίες του πάνω στην Ανάλυση Παραγόντων. Ο Spearman ήρθε σε αντιπαράθεση με τον Pearson σχετικά με το αν ο δείκτης ρ ήταν το ίδιο χρήσιμος με τον δείκτη r .

Σε γενικές γραμμές ο r είναι ένα κλάσμα που έχει στον αριθμητή τη συνδιακύμανση των μεταβλητών, δηλαδή το άθροισμα των γινομένων των αποστάσεων των αντίστοιχων τιμών των δύο μεταβλητών από τον μέσο όρο καθεμιάς και στον παρονομαστή το γινόμενο των διακυμάνσεων, ώστε να παίρνει τιμές από -1 ως +1. Ο δείκτης r είναι από τους πιο σπουδαίους δείκτες συνάφειας και συνδέεται με την ανάλυση διακύμανσης και τη διαφορά μέσων όρων των μεταβλητών. Ονομάζεται δε και «γινόμενο των ροπών» διότι βασίζεται στα γινόμενα των αποστάσεων των τιμών από τον μέσο όρο τους («ροπή» είναι μια έννοια στη Νευτώνεια φυσική που ισούται με δύναμη επί απόσταση). Ο δείκτης ρ του Spearman από την άλλη μεριά έχει στον αριθμητή το άθροισμα των τετραγώνων των διαφορών των τακτικών τιμών των δύο μεταβλητών («τακτικές» γιατί έχουν καταταχθεί σε πρώτη, δεύτερη, τρίτη κλπ., ανεξάρτητα από την τιμή τους) και στον παρονομαστή το ανώτατο δυνατό άθροισμα των διαφορών των τακτικών τιμών. Το όλο κλάσμα αφαιρείται από τη μονάδα, αφού πρώτα ο αριθμητής του πολλαπλασιαστεί με 6. Το «6» έχει να κάνει με την ανάγκη το κλάσμα να παίρνει τιμές μεταξύ του -1 και του 1. Ο παρονομαστής του ρ είναι $n(n^2-1)/3$ και η ευθεία που περνάει από τα σημεία $(0, 1)$ και $(n(n^2-1)/3, -1)$ έχει κλίση $-6/n(n^2-1)$ και αποτέμνουσα ίση με 1.

Ο Πίνακας 2 στη σελίδα 296 της *Ψυχολογίας* περιέχει μέσους όρους και τυπικές αποκλίσεις για αγόρια και κορίτσια σε τάξεις του Δημοτικού σχολείου. Με «N» σημειώνεται το μέγεθος του δείγματος, ενώ οι συγγραφείς παρουσιάζουν τα εκατοστημόρια των κατανομών των σωστών απαντήσεων. Το p-οστό εκατοστημόριο μιας κατανομής είναι εκείνη η τιμή, η οποία, όταν οι τιμές διαταχθούν σε αύξουσα σειρά, έχει από αριστερά της το p% των δεδομένων και από δεξιά της το υπόλοιπο (100-p)%. Το πεντηκοστό τεταρτημόριο είναι η διάμεσος, η τιμή που έχει τις μισές τιμές μεγαλύτερες από αυτήν και τις μισές τιμές μικρότερη από αυτήν. Με τον Πίνακα 2 οι συγγραφείς κατορθώνουν να μας δώσουν μια εικόνα των κατανομών.



Σχήμα 2. Διδιάστατη κανονική κατανομή με δύο κατηγορίες ανά μεταβλητή.

Οι στατιστικά «σημαντικές διαφορές μεταξύ των δύο φύλων», όπως γράφουν οι συγγραφείς στη σελίδα 296 εξακριβώθηκαν με το κριτήριο F. Το εν λόγω κριτήριο λέγεται έτσι επειδή ακολουθεί την κατανομή F, η οποία με τη σειρά της ονομάστηκε έτσι από το αρχικό του επιθέτου του Sir Ronald Fisher που το πρωτοκατασκεύασε για τον έλεγχο κυρίως της ισότητας των διακυμάνσεων. Εδώ, όμως, το F τεστ χρησιμοποιείται από τους συγγραφείς ως εναλλακτικό του t τεστ για να διερευνηθούν οι διαφορές ανάμεσα στα αγόρια και στα κορίτσια. Το t τεστ πήρε το όνομά του από τον «φοιτητή» (student), όπως υπέγραψε το 1908 τη σχετική εργασία του ο χημικός και στατιστικός William Sealy Gosset στο περιοδικό *Biometrika* που διήυθνε ο Karl Pearson. Ο Gosset εργαζόταν για τη εταιρεία μύρας Guinness στο Δουβλίνο. Η εν λόγω εταιρεία

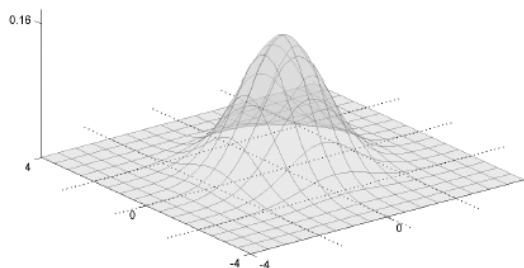
προσλάμβανε με καλό μισθό τους κορυφαίους απόφοιτους της Οξφόρδης και του Κέμπριτζ, αλλά τους απαγόρευε να δημοσιεύουν επιστημονικά κείμενα. Έτσι ο Gosset δημοσίευσε με το ψευδώνυμο «φοιτητής».

Με το t τεστ ελέγχουμε τη στατιστική σημαντικότητα της διαφοράς δύο μέσων όρων. Το t τεστ είναι κι αυτό ένα κλάσμα, το οποίο έχει στον αριθμητή τη διαφορά των μέσων όρων και στον παρονομαστή ένα κάποιο μέτρο της κοινής διακύμανσης ή των διαφορετικών διακυμάνσεων των δύο γκρουπ. Όσο πιο μεγάλη είναι η διαφορά των μέσων όρων και όσο πιο κοντά στον δικό τους η καθεμιά μέσο όρο είναι οι τιμές των δύο γκρουπ που συγκρίνουμε, τόσο μεγαλύτερο είναι το t τεστ και είναι πιθανότερη η διάψευση της μηδενικής υπόθεσης. Το F τεστ, από την άλλη μεριά, είναι και αυτό ένα κλάσμα, το οποίο όμως έχει στον αριθμητή τη διακύμανση μεταξύ των μέσων όρων όλων των γκρουπ (συνήθως περισσότερων των δύο) και στον παρονομαστή τη συνολική διακύμανση όλων των τιμών ανεξαρτήτως γκρουπ. Αν η διακύμανση στον αριθμητή είναι μεγάλη σε σχέση με τη διακύμανση στον παρονομαστή, η τιμή του F είναι πολύ πιθανό να είναι στατιστικώς σημαντική. Οι αγκύλες δεξιά από το γράμμα F (βλ. Σιδερίδης κ. ά. 2008: 296) περικλείουν τους βαθμούς ελευθερίας και είναι σημαντικοί γιατί η δειγματοληπτική κατανομή του F, στη βάση της οποίας ελέγχουμε την τιμή που βρήκαμε και απορρίπτουμε ή αποτυγχάνουμε να απορρίψουμε τη μηδενική υπόθεση, είναι διαφορετική για διαφορετικούς βαθμούς ελευθερίας. Όπως είδαμε στην σελίδα **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**, οι βαθμοί ελευθερίας είναι πόσα στοιχεία σε μια στατιστική ανάλυση είναι ελεύθερα να διαφέρουν. Στην Ανάλυση Διακύμανσης (ANOVA) έχουμε δύο

βαθμούς ελευθερίας. Ο πρώτος, έστω k , εξαρτάται από τον ερευνητικό σχεδιασμό. Ο δεύτερος είναι στην ουσία το μέγεθος του δείγματος πλην τους βαθμούς ελευθερίας του ερευνητικού σχεδιασμού. Γενικά σε σύνθετους ερευνητικούς σχεδιασμούς ανάλυσης διακύμανσης τα γκρουπ για τον υπολογισμό της διακύμανσης στον αριθμητή του F είναι τόσα όσες είναι οι δυνατές επιμέρους περιπτώσεις. Στη δική μας περίπτωση έχουμε συνολικά στην έρευνα $k=10$ περιπτώσεις, 2 για το φύλο $\times 5$ για κάθε μία από τις πέντε σχολικές τάξεις. Στον παρονομαστή του κριτηρίου F , δηλαδή στη συνολική διακύμανση, οι βαθμοί ελευθερίας είναι $n - k$, όπου n το μέγεθος του δείγματος. Στην περίπτωση που μελετάμε εδώ έχουμε μόνο δύο μέσους όρους και έτσι στον αριθμητή έχουμε $2-1=1$ βαθμούς ελευθερίας. Έτσι η διακύμανση στον αριθμητή είναι απλώς η διαφορά των δύο μέσων όρων. Στην περίπτωση αυτή η τιμή του κριτηρίου F είναι η τιμή του κριτηρίου t υψωμένη στο τετράγωνο. Οι συγγραφείς χρησιμοποίησαν το κριτήριο F και έλεγξαν επίσης την αλληλεπίδραση μεταξύ φύλου και τάξης. Και αυτή η αλληλεπίδραση βρέθηκε μη στατιστικώς σημαντική.

ΑΝΙΧΝΕΥΣΗ ΑΠΛΩΝ ΔΟΜΩΝ

Στην επόμενη ενότητα οι συγγραφείς (Σιδερίδης κ.ά. 2008) προχωρούν στην «ανίχνευση απλών δομών ή την ανίχνευση της κατασκευαστικής δομικής εγκυρότητας του ψυχομετρικού εργαλείου μέσω της Ανάλυσης Κυρίων Συνιστωσών (Principal Components Analysis) που είδαμε κι εμείς. Η ανάλυση αυτή ανήκει στην κλασική θεωρία των μετρήσεων και η χρήση της εδώ είναι ένα παράδειγμα του πώς οι δύο θεωρίες, η κλασική και η νέα, μπορούν να συνεργάζονται. Η χρήση της Ανάλυσης Κύριων Συνιστωσών χρησιμοποιείται εδώ διότι στο συγκεκριμένο τεστ υποτίθεται ότι μόνο μία θεωρητική κατασκευή «ευθύνεται» για τη διακύμανση στην κατανομή των σωστών απαντήσεων των μαθητών. Αυτή η κατασκευή πρέπει είναι η «ορθογραφική ικανότητα». Τυχόν ύπαρξη κι άλλων λανθάνουσών μεταβλητών περα από την ορθογραφική ικανότητα θα έθετε σε κίνδυνο την αξιοπιστία των μετρήσεων.

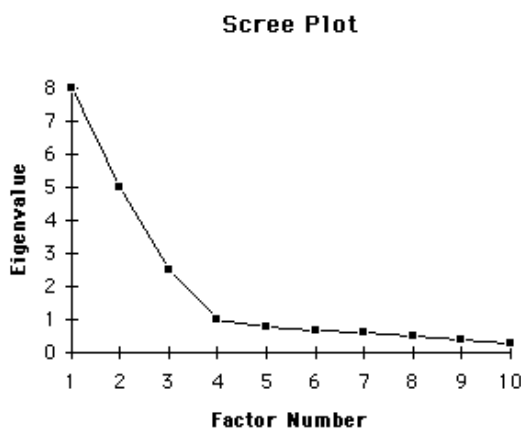


Σχήμα 3. Διδιάστατη κανονική κατανομή με τέσσερις τεχνητές κατηγορίες ανά μεταβλητή.

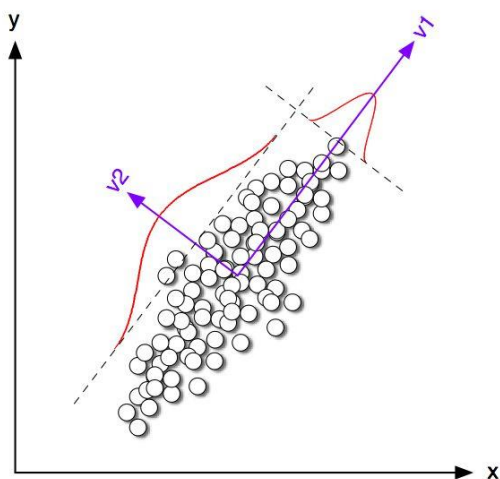
Η ύπαρξη μιας και μόνης διάστασης είναι πολύ σημαντικό ζήτημα στην κατασκευή ενός τεστ. Υποτίθεται ότι η κλίμακα μετράει μία και μόνη λανθάνουσα μεταβλητή. Ένας τρόπος να το δει κανείς αυτό είναι να μελετήσει τη δομή της συνδιακύμανσης ανάμεσα στα ερωτήματα του τεστ. Οι συγγραφείς αναφέρουν στη σελίδα 297 ότι ένας τρόπος για να ελεγχθεί αυτό θα ήταν μέσω του λογισμικού EQS 6.1 και της

χρήσης πολυχωρικών συντελεστών συνάφειας. Το λογισμικό EQS επικεντρώνεται

στην κατασκευή δομικών μοντέλων. Τα δομικά μοντέλα είναι ένα σύνολο τεχνικών για την διερεύνηση σχέσεων μεταξύ λανθανουσών και παρατηρούμενων μεταβλητών. Οι συντελεστές συνάφειας αποτελούν τη βάση για την κατασκευή των εν λόγω μοντέλων. Στην έρευνα που διαβάζουμε παρατηρούμενες μεταβλητές είναι η ορθή ή η λανθασμένη γραφή των 60 λέξεων ενώ η μόνη υποτιθέμενη λανθάνουσα μεταβλητή είναι η ορθογραφική ικανότητα των μαθητών. Μέσω της τελευταίας εμφανίζεται η ορθή γραφή των λέξεων¹ και «εξηγείται» η διακύμανση μεταξύ ορθής και λανθασμένης γραφής. Οι 60 λέξεις του τεστ θα σχημάτιζαν συνολικά 1770 ζευγάρια αν παίρνονταν ανά δύο. Κάθε ζευγάρι θα σχημάτιζε έναν «χώρο» με τέσσερις περιπτώσεις: (α) και οι δύο λέξεις να είναι γραμμένες ορθά, (β) και οι δύο λέξεις να είναι γραμμένες λάθος, (γ) σωστή η πρώτη λέξη και λάθος η δεύτερη, και (δ) σωστή η δεύτερη λέξη και λάθος η πρώτη.



Σχήμα 5. Scree test για την απόφαση του πόσες κύριες συνιστώσες θα διατηρηθούν στο μοντέλο.



Σχήμα 4. Καρτεσιανό σύστημα συντεταγμένων και δύο κύριες συνιστώσες στα δεδομένα.

να είναι γραμμένες ορθά, (β) και οι δύο λέξεις να είναι γραμμένες λάθος, (γ) σωστή η πρώτη λέξη και λάθος η δεύτερη, και (δ) σωστή η δεύτερη λέξη και λάθος η πρώτη.

Ο συντελεστής συνάφειας μεταξύ της επιτυχίας ή αποτυχίας στην ορθή γραφή δύο λέξεων λαμβάνει υπόψη του το γεγονός ότι ο διαχωρισμός «σωστό – λάθος» είναι στην ουσία η αποτύπωση ενός παράγοντα που ακολουθεί την κανονική κατανομή. Αυτός ο παράγοντας είναι φυσικά η ορθογραφική ικανότητα. Έτσι, οι δύο λέξεις, ως ζευγάρι πλέον, ακολουθούν τη διδιάστατη κανονική κατανομή, η οποία εμφανίζεται στον τριδιάστατο χώρο στο Σχήμα 2 και στο Σχήμα 3. Γενικά, οι «πολυχωρικοί» συντελεστές λαμβάνουν υπόψη τους την τεχνητή κατάτμηση

¹Πηγές για τη λογική των δομικών μοντέλων υπάρχουν στο διαδίκτυο (βλ. <http://davidakenny.net/cm/causal.htm>), καθώς και στο περιοδικό Structural Equation Modeling του οίκου Taylor and Francis.

μεταβλητών (κανονικές κατανομές) σε δύο η περισσότερες κατηγορίες. Στο Σχήμα 3 εμφανίζεται η διδιάστατη κανονική κατανομή και οι τέσσερις κατηγορίες κάθε μιας μεταβλητής. Για τον υπολογισμό των πολυχωρικών μεταβλητών έχουν προταθεί πολλοί αλγόριθμοι, ενώ ειδικά για τους τετραχωρικούς έχει υποστηριχθεί ότι υπό προϋποθέσεις θα μπορούσε να χρησιμοποιηθεί αντί για αυτούς κάποιος απλός συντελεστής συνάφειας για κατηγορικές μεταβλητές με δύο όμως κατηγορίες έκαστη (Kolenikov & Angeles, 2004). Ο υπολογισμός του τετραχωρικού συντελεστή συνάφειας έχει συζητηθεί ευρέως στην στατιστική βιβλιογραφία (βλ. Divgi, 1979· Manal, 2002· Bonet & Price, 2005). Η πιο κοινή μέθοδος για την προσέγγισή του είναι μέσω του συντελεστή ϕ . Στην περίπτωση αυτή ο τετραχωρικός συντελεστής r_t είναι το συνημίτονο της ποσότητας $\pi \cdot \nu\beta\gamma / \nu\alpha\delta + \nu\beta\gamma$, όπου $\alpha, \beta, \gamma, \delta$, είναι τα ποσοστά των μετρήσεων στα κελιά ενός πίνακα 2 επί 2 που σχηματίζονται από το πρώτο, το δεύτερο, το τρίτο και το τέταρτο τεταρτημόριο ενός κύκλου που ορίζεται από τις γραμμές διχοτόμησης των κατανομών δύο συσχετιζόμενων ερωτημάτων (Bouvier, Perry, & Michael, 1954).

SCREE ΤΕΣΤ

Άλλο ένα κριτήριο για να δει κανείς αν ένας και μόνο παράγοντας «εξηγεί» τη διακύμανση σε ένα σύνολο ερωτημάτων είναι το Scree τεστ, ένα οπτικό τεστ, το οποίο παράγεται κατά τη διαδικασία της διερευνητικής ανάλυσης παραγόντων ή ορθότερα της διερευνητικής ανάλυσης κυρίων συνιστωσών. Το εν λόγω τεστ έχει πάρει το όνομά του από το σχήμα που έχουν τα αμμώδη υλικά όταν πέφτουν από ψηλά και συσσωρεύονται σε ένα “scree” ή στα ελληνικά σε ένα «κόρημα».

Κεντρική ιδέα της ανάλυσης συνιστωσών είναι ότι η συνολική διακύμανση σε ένα σύνολο παρατηρήσεων μπορεί να παρουσιαστεί με τη βοήθεια συνιστωσών, οι οποίες είναι λιγότερες αριθμό από τις αρχικές μεταβλητές και «εξηγούν» ένα σημαντικό μέρος της διακύμανσης. Για παράδειγμα, στο Σχήμα 4 βλέπουμε ένα σύνολο παρατηρήσεων σε ένα διάγραμμα διασποράς. Είδαμε στη σελίδα 28 πώς υπολογίζουμε τις ιδιοτιμές και τα ιδιοδιανύσματα ενός τετραγωνικού πίνακα. Στην παρούσα έρευνα ο τετραγωνικός πίνακας έχει διαστάσεις 60×60 αφού τόσες είναι οι λέξεις. Στο Σχήμα 4 η δομή των δεδομένων που περιγράφεται στο καρτεσιανό σύστημα αξόνων y και x , μπορεί να παρουσιαστεί με τη βοήθεια των συνιστωσών v_1 και v_2 . Τα ιδιοδιανύσματα των v_1 και v_2 είναι η διεύθυνση των νέων αξόνων αναφοράς ενώ η ιδιοτιμή του καθενός είναι το μέτρο τους. Η ιδιοτιμή του v_1 είναι πολύ μεγαλύτερη από εκείνη του v_2 γιατί η πρώτη «εξηγεί» πολύ μεγαλύτερη διακύμανση. Αν επιλέγαμε να μην είχαμε τον v_2 , τα δεδομένα θα

μπορούσαν να παρουσιαστούν ως προβολές στον v_1 , με κάποια θυσία ως προς τη συνολική διακύμανση που για να παρουσιαστεί πλήρως απαιτεί και τους δύο άξονες. Στην περίπτωση αυτή όλες οι τιμές θα προβάλλονταν στον v_1 , ενώ η διακύμανση στον v_2 θα είχε χαθεί. Αν αυτό το φανταστούμε όχι σε δύο αλλά σε n

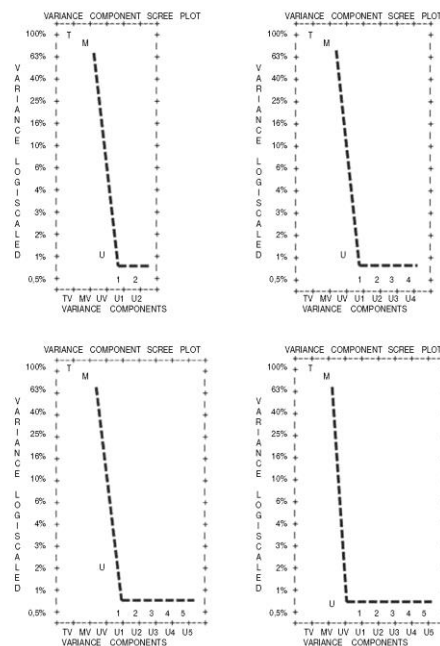
διαστάσεις, παίρνουμε μια ιδέα για τη χρησιμότητα της «Κύριων Συνιστωσών» ως μεθόδου μείωσης των διαστάσεων σε έναν χώρο διαστάσεων n .

Το Scree τεστ, σύμφωνα με την εμπειρική πρακτική, προτείνει να κρατάμε στην ανάλυση μία λιγότερη συνιστώσα από όσες

υπερβαίνουν το «μανίκι» του σχήματος, το σημείο, δηλαδή στο οποίο το σχήμα οριζοντιώνεται. Στο παράδειγμα (Σχήμα 5) η

οριζοντίωση διαφαίνεται στην τέταρτη συνιστώσα πράγμα που σημαίνει ότι κρατάμε τρεις παράγοντες (για την ακρίβεια «συνιστώσες»).

Στην περίπτωση της έρευνας των Σιδερίδη κ. ά. (2008), όμως, τα διαγράμματα κορήματος που παρουσιάζονται στη σελίδα 297 δεν αναφέρονται στην αρχική διακύμανση των 60 λέξεων και στους πολυχρωικούς (ενδεχομένως) συντελεστές τους όπως θα απαιτούσε η Ανάλυση Κύριων Συνιστωσών, αλλά στην απομένουσα διακύμανση, αφού πρώτα έχει εφαρμοστεί στις λέξεις του τεστ το μοντέλο Rasch για το οποίο θα μιλήσουμε σε λίγο. Η διακύμανση στον κάθετο άξονα είναι η εναπομένουσα διακύμανση, η οποία δίνεται από το Winsteps ως ποσοστό αλλά σε



Εικόνα 1
Διαγράμματα Scree για τη Β' (πάνω αριστερά), Γ' (πάνω δεξιά), Δ' τάξη (κάτω αριστερά), και για το συνολικό δείγμα (κάτω δεξιά)

EQS 6.1. Bentler, 2004) ή (β) με την εφαρμογή του μοντέλου Rasch. Δεδομένου ότι οι κύριοι υποστηρικτές της ανάλυσης συνιστωσών προτείνουν τη δεύτερη μέθοδο ως πιο κατάλληλη για διχοτομικά, κατηγορικά δεδομένα (Bentler, in press, Joreskog & Moustaki, 2001. Moustaki, 2001), χρησιμοποιήθηκε η δεύτερη μέθοδος. Με-

ταξύ των κριτηρίων ανίχνευσης διαστάσεων το επικρατέστερο είναι αυτό της οπτικής ανάλυσης των σχημάτων Scree (Gorsuch, 1983). Εντοπίζοντας τις ευθείες που σχηματίζονται από δύο ή περισσότερες ιδιοτιμές, ο ερευνητής μπορεί να αξιολογήσει την ύπαρξη απλών δομών. Οι αναλύσεις πραγματοποιήθηκαν ξεχωριστά για τη Β',

λογαριθμική κλίμακα, πράγμα που σημαίνει οι διαφορές στο ανώτερο τμήμα του κάθετου άξονα είναι πολύ πιο σημαντικές από τις διαφορές στο κάτω μέρος του ίδιου άξονα. Στα Scree test των Σιδερίδη κ. ά. (2008) βλέπουμε τη συνολική διακύμανση, η οποία σημειώνεται στο εκατό τοις εκατό με το γράμμα “T” (από το “Total”). Στα σχήματα φαίνεται επίσης για κάθε σχολική τάξη: (α) η

298 ♦ Γεώργιος Δ. Σιδερίδης, Αγγελική Μουζάκη, Αθανάσιος Προτόπαπας & Παναγιώτης Σίμος

Πίνακας 3
Σταθμισμένοι συντελεστές από την ανάλυση Κυρίων Συνιστωσών με τη χρήση του Μοντέλου Rasch για την αξιολόγηση της ύπαρξης δεύτερης διάστασης στη δοκιμασία της ορθογραφίας. Μικροί συντελεστές υποστηρίζουν την ύπαρξη μιας μόνο διάστασης.

A/A	Λέξη	B'	Γ'	Δ'	Σύνολο	A/A	Λέξη	B'	Γ'	Δ'	Σύνολο
1	από	0,02	0,09	0,00	0,04	31	αυτοκίνητο	-0,29	0,03	-0,15	-0,03
2	έλα	0,07	-0,05	-0,22	-0,03	32	παράδειγμα	-0,17	-0,05	0,25	0,16
3	και	-0,01	0,08	0,00	0,05	33	φωτισμένος	-0,04	0,01	-0,29	-0,33
4	είναι	0,04	-0,28	-0,07	-0,14	34	άδεια	-0,25	-0,33	-0,15	-0,15
5	πετάνε	-0,06	0,08	-0,06	0,08	35	ζεσταίνει	0,17	-0,01	0,01	-0,04
6	άλογο	-0,02	-0,07	0,01	0,01	36	ονειρεύομαι	0,26	-0,02	0,01	-0,02
7	φίλος	0,13	0,04	-0,25	-0,14	37	αλλιώτικος	-0,11	-0,02	-0,25	-0,26
8	ξέρω	0,09	0,02	0,03	-0,09	38	δανειζώ	-0,22	0,05	-0,11	-0,11
9	έτσι	0,09	0,02	-0,11	-0,09	39	μόλυνση	-0,11	0,13	-0,08	-0,01
10	όμως	-0,08	0,03	0,15	0,05	40	πρακτορείο	0,08	-0,02	-0,07	-0,11
11	ούτε	0,06	-0,06	0,10	0,09	41	πετρέλαιο	0,74	0,07	0,09	0,14
12	λέω	-0,05	0,00	-0,03	-0,02	42	οικονομικότερος	0,20	0,00	-0,31	-0,18
13	αυτός	0,08	-0,05	0,07	-0,02	43	υποχρεωμένος	0,20	0,19	0,08	0,14
14	τραπέζι	0,10	0,00	-0,29	-0,23	44	αναχώρηση	0,05	0,01	-0,12	-0,05
15	ποτιζώ	0,18	0,06	0,20	0,09	45	χαρακτηριστικό	0,03	-0,09	0,18	-0,05
16	πηγή	-0,21	0,04	-0,04	0,08	46	ξεφυλλίζοντας	0,11	-0,18	-0,03	-0,20
17	φιλί	-0,19	0,01	0,03	-0,07	47	κυνηγητό	0,73	0,28	0,04	0,11
18	φυτό	-0,18	0,04	0,09	-0,09	48	ηθοποιός	0,70	0,36	0,17	0,26
19	δύχτυ	0,01	-0,17	-0,04	-0,10	49	διευθυντής	-0,22	0,28	0,34	0,48
20	χρήματα	0,02	-0,03	-0,31	-0,01	50	αντικείμενο	0,41	0,12	0,60	0,57
21	πέτρινο	-0,09	0,06	0,26	0,14	51	επιχείρηση	0,27	-0,04	0,25	0,20
22	ζωγραφική	0,03	-0,01	-0,05	-0,09	52	εκατομμύριο	0,29	0,33	0,45	0,58
23	πηγαίνω	0,10	-0,14	0,06	-0,09	53	ειδοποιήθηκε	0,000	0,50	-0,14	0,17
24	μεγαλώνω	0,11	-0,79	-0,08	-0,32	54	αισθήσεις	0,000	0,41	0,28	0,34
25	σοφοί	-0,32	-0,51	-0,08	-0,25	55	νοημοσύνη	0,000	-0,06	0,43	0,32
26	δωρεάν	0,02	-0,38	-0,36	-0,30	56	αποχαιρέτησα	0,000	0,15	0,39	0,26
27	παιχτών	-0,14	-0,05	-0,30	-0,20	57	χειμαρρος	0,000	0,00	-0,07	0,03
28	ταμείο	-0,08	-0,32	-0,20	-0,16	58	εγχειριστεί	0,000	0,81	0,02	0,24
29	χώρισα	-0,24	0,03	-0,06	-0,03	59	ματαιώνεται	0,000	0,18	0,11	0,20
30	επιβάτης	0,22	-0,17	0,17	0,03	60	αποδεικνύω	0,000	0,000	0,000	0,000

Σημείωση: Σε σκίαση παρουσιάζονται λέξεις για τις οποίες φαίνεται να αποδίδεται διακύμανση σε δεύτερο παράγοντα.

Γ', και Δ' τάξη με βάση τα δεδομένα της πρώτης μέτρησης, καθώς διαπιστώθηκαν σημαντικές αυξητικές τάσεις της συνολικής επίδοσης. Στα διαγράμματα Scree, που παρατίθενται στην Εικόνα 1, φαίνεται ότι οι μεγαλύτερες ιδιοτιμές (πρώτη

και δεύτερη) εμπίπτουν σε μια ευθεία γραμμή, ενώ οι υπόλοιπες «καταλήγουν» σε μια ασυμπτωτική ευθεία και εξηγούν ποσοστά διακύμανσης κάτω από 1%. Συγκεκριμένα τα ποσοστά διακύμανσης που αποδίδονται στον πρώτο πα-

επεξηγούμενη διακύμανση μέσω του μοντέλου Rasch (σημειώνεται με “M” από το “Modeled”), δηλαδή η διακύμανση μεταξύ των λέξεων αν σε αυτές υπήρχε μόνο μία διάσταση όπως είναι η προϋπόθεση του μοντέλου Rasch. Φαίνεται επίσης και η μη επεξηγούμενη διακύμανση από το μοντέλο Rasch, η οποία σημειώνεται ως “U” από το αρχικό της λέξης “Unexplained”).

Απειλή για την μονοδιαστατικότητα του τεστ δεν θα αποτελούσε το ποσοστό της επεξηγούμενης προς τη μη επεξηγούμενη διακύμανση, αλλά και η ύπαρξη

συνιστώσών με ποσοστό διακύμανσης πάνω από 2% της συνολικής, αφού χρειάζονται τουλάχιστον 2 ερωτήματα ώστε η συνιστώσα να κατοχυρωθεί ως «διάσταση».. Σε καμία σχολική τάξη από τις τέσσερις που παρουσιάζονται στην Εικόνα 1 του κειμένου των Σιδερίδη κ. ά. (2008) οι συνιστώσες U1, U2 ... δεν υπερέβησαν το 2%, πράγμα που σημαίνει ότι η προϋπόθεση της μονοδιαστατικότητας στην έρευνα που μελετάμε ισχύει.

Στον Πίνακα 3 στη σελίδα 298 της εργασίας τους οι ερευνητές (Σιδερίδης κ. ά., 3008) παρουσιάζουν τους συντελεστές συνάφειας των 60 λέξεων με ένα υποτιθέμενο δεύτερο παράγοντα, πέρα από την ορθογραφική ικανότητα. Φαίνεται ότι μόνο λίγες λέξεις φανερώνουν μια τέτοια πιθανή δομή. Για παράδειγμα, η λέξη «κυνηγητό» μοιάζει να επηρεάζεται και από έναν δεύτερο παράγοντα, αφού ο συντελεστής συνάφειας με έναν τέτοιο πιθανό παράγοντα είναι για την Β' τάξη 0,73. Η λέξη «ηθοποιός» επίσης επηρεάζεται πιθανώς και από έναν δεύτερο παράγοντα, αφού στη Β και τη Γ' τάξη οι συντελεστές συνάφειας είναι 0,70 και 0,36 αντιστοίχως (θα έπρεπε να είναι κοντά στο μηδέν σε όλες τις λέξεις). Οι συγγραφείς στηρίζουν βιβλιογραφικά τη θέση τους περί ύπαρξης ενός και μόνο παράγοντα ενώ αναφέρουν ότι η λέξη «αποδεικνύω» θεωρείται ως «σταθερά» για τον παράγοντα της ορθογραφικής ικανότητας, αφού η συνάφειά του με κάποιον υποτιθέμενο παράγοντα πέραν της ορθογραφικής ικανότητας είναι ουσιαστικά μηδέν.

ΤΟ ΜΟΝΤΕΛΟ RASCH

Στη συνέχεια οι συγγραφείς προχωρούν στην κύρια πλευρά της ανάλυσής τους, και πιο συγκεκριμένα στην ανάλυση του βαθμού δυσκολίας των λέξεων μέσω του μοντέλου Rasch.

Στη σχετική ενότητα (Σιδερίδης κ. ά.: 299) εξηγούν τη λογική της ανάλυσης κατά Rasch. Σε γενικές γραμμές, εξηγούμε εμείς, τα μοντέλα Rasch υπολογίζουν την πιθανότητα να απαντήσει κάποιος μαθητής j σωστά σε μια ερώτηση i (εδώ να γράψει σωστά την λέξη), με δεδομένη την ορθογραφική του ικανότητα αλλά και τη δυσκολία της συγκεκριμένης λέξης. Αυτό υπολογίζεται από το πώς έχουν απαντήσει οι άλλοι συμμετέχοντες μαθητές στις ίδιες ερωτήσεις (λέξεις). Στην ιδανική περίπτωση οι μαθητές απαντούν σωστά μόνο σε ερωτήσεις που

ράγοντα ήταν 98,8% για τη Β' τάξη, 98,7% για την Γ' τάξη, 98,2% για τη Δ' τάξη, και 99,5% για το συνολικό δείγμα. Σύμφωνα με τον Linacre (1994), η πρώτη διάσταση υποστηρίζεται όταν σε αυτήν κυμαίνεται πάνω από 60% της συνολικής διακύμανσης. Αν η δεύτερη διάσταση δεν ερμηνεύει τουλάχιστον 5% της συνολικής θεωρείται ότι η ύπαρξή της δεν υποστηρίζεται από τα δεδομένα. Συνεπώς, αυτά τα αποτελέσματα δείχνουν ότι υπάρχει μια κύρια διάσταση ορθογραφικής επίδοσης, ενώ τα υπολείμματα από αυτή τη διάσταση αποδίδονται (αναγκαστικά) σε ένα δεύτερο παράγοντα ο οποίος δεν είναι απαραίτητα ερμηνεύσιμος (επειδή εξηγεί πολύ μικρά ποσοστά διακύμανσης). Συμπερασματικά, το μοντέλο ενός παράγοντα ταιριάζει στα δεδομένα για όλες τις τάξεις και υποστηρίζει τη μονοδιάστατη δομή της ορθογραφικής δεξιότητας κατά τη γραφή μεμονωμένων λέξεων.

Ο Πίνακας 3 δείχνει τους σταθμισμένους συντελεστές που προέκυψαν αφού αφαιρέθηκε η συνεισφορά του πρώτου παράγοντα, δηλαδή από την ανάλυση των «υπολειμμάτων» ή σφαλμάτων, αφού αφαιρέθηκε η διακύμανση που οφειλόταν στην πρώτη διάσταση. Για να επιβεβαιωθεί ότι δεν υπάρχει δεύτερη διάσταση θα πρέπει τα υπολείμματα αυτά να είναι αποσυσχετισμένα και αρκετά «μικρά» (κάτω από 0,40: Gorsuch, 1983). Αυτή η ανάλυση είναι καθοριστική για την αξιολόγηση διατάσεων πέρα από την πρώτη που προέκυψε από την εφαρμογή του μοντέλου Rasch. Όπως δείχνει ο Πίνακας 3, οι περισσότερες λέξεις είχαν, όπως ήταν επιθυμητό, μικρή έως και μηδαμινή φόρτιση στο δεύτερο παράγοντα. Ελάχιστες λέξεις με κοινό χαρακτηριστικό τους τον υψηλό βαθμό δυσκολίας παρουσίασαν υψηλότερες φορτίσεις. Συνολικά όμως τα ποσοστά της επιπλέον αποδιδόμενης διακύμανσης από το δεύτερο παράγοντα ήταν πολύ μικρά (λιγότερο από 2,5%). Δεδομένου του διερευνητικού χαρακτήρα της μελέτης αυτής (δηλ. της προσπάθειας δημιουργίας μιας δοκιμασίας ορθογραφικής ικανότητας) είναι σημαντικό να εξεταστεί μελλοντικά η συνεισφορά της τελευταίας λέξης («αποδεικνύω»), η οποία φαίνεται να αποτελεί «σταθερά»

για το μοντέλο Rasch, δηλαδή μάλλον είναι μη υπολογίσιμη δυσκολίας για τις ηλικίες του δείγματος (τάξεις Β' - Δ').

Προσδιορισμός της σχετικής δυσκολίας των λέξεων της δοκιμασίας

Ο σκοπός αυτής της ανάλυσης ήταν (α) να αξιολογήσει τη συνεισφορά κάθε επιμέρους λέξης στη συνολική πληροφόρηση που παρέχει η δοκιμασία, και (β) να υπολογίσει την καταλληλότερη σειρά των λέξεων με βάση το σχετικό βαθμό δυσκολίας τους. Παρακάτω, αρχικά παρουσιάζεται το μοντέλο Rasch και στη συνέχεια παρατίθενται τα αποτελέσματα από την εφαρμογή του στην παρούσα μελέτη.

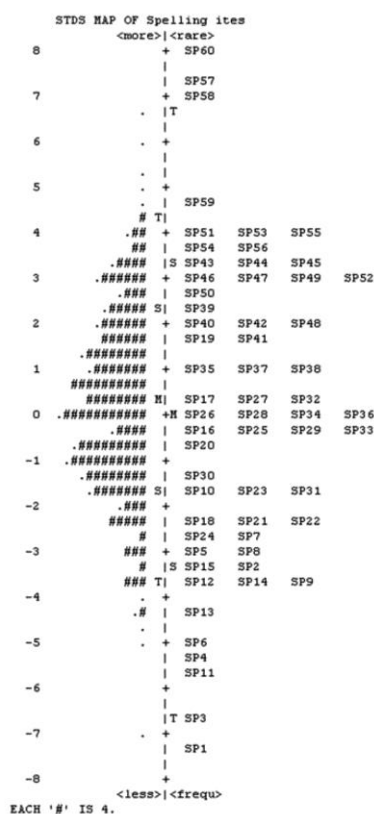
Το μοντέλο Rasch (1980) αναπτύχθηκε με το σκεπτικό ότι ένα μη γραμμικό μοντέλο μπορεί να είναι καταλληλότερο για την αξιολόγηση της επίδοσης και της συμπεριφορά ατόμων αλλά και των μερών ενός τεστ (βλέπε Smith & Smith, 2004). Το παραδοσιακό μοντέλο Rasch αξιολογεί την πιθανότητα ενός ατόμου να απαντήσει σε μια κλίμακα (ή μια επιμέρους ερώτηση ενός τεστ) ως συνάρτηση δύο παραμέτρων: (α) της ικανότητας του ατόμου (B) όπως μετρείται από τη συνολική του βαθμολογία στο τεστ, και (β) της ειδικής δυσκολίας (D) της συγκεκριμένης κλίμακας (ή ερώτησης). Η εξίσωση που περιγράφει αυτό το μοντέλο για διχοτομικά δεδομένα (του τύπου σωστό/λάθος) είναι η ακόλουθη (Bond & Fox, 2001. Rasch, 1980. Wright & Stone, 1979):

$$P_{ni}(X_{ni} = 1/B_n, D_i) = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}} \quad (1)$$

Η παράμετρος $P_{ni}(X_{ni} = 1/B_n, D_i)$ εκφράζει την πιθανότητα το άτομο n να απαντήσει σωστά στην ερώτηση i , δεδομένης της ικανότητάς του (B) και της δυσκολίας της ερώτησης (D), μετρημένες σε λογαριθμικές μονάδες (logit). Η παράμετρος e είναι η βάση των φυσικών λογαρίθμων (2,71828...). Σύμφωνα με την εξίσωση (1) η πιθανότητα να απαντήσει σωστά την ερώτηση ένα άτομο εξαρτάται από τη διαφορά μεταξύ της ικα-

αντικατοπτρίζουν το επίπεδο της ικανότητάς τους, καθώς και σε όλες τις ευκολότερες, αλλά όχι στις δυσκολότερες σε σχέση με το επίπεδο της ικανότητάς τους. Στην περίπτωση αυτή οι ερωτήσεις λέγεται ότι σχηματίζουν μια κλίμακα Guttman, στην οποία η ικανότητα προχωράει βήμα-βήμα με την δυσκολία. Για αυτό και το σχετικό λογισμικό για τα μοντέλα Rasch έχει τη φράση “steps” στο λογότυπό του και ονομάζεται “Winsteps”. Κάθε μαθητής, ανάλογα με την ικανότητά του, κάνει «βήματα» σωστών απαντήσεων. Το μοντέλο Rasch μοντελοποιεί τη πιθανότητα ώστε οι μαθητές να κάνουν αυτά τα «βήματα» που κάνουν ή που πρόκειται να κάνουν. Ο τρόπος με τον οποίο τα μοντέλα αυτά χειρίζονται την πιθανότητα αναφέρεται στις παραγράφους που ακολουθούν.

Γενικά, ο πιο συνηθισμένος ορισμός της πιθανότητας p είναι ότι αυτή εκφράζεται με ένα κλάσμα που έχει στον αριθμητή κάποιο επιθυμητό ενδεχόμενο και στον παρονομαστή όλα τα πιθανά ενδεχόμενα μαζί. Η πιθανότητα παίρνει τιμές στο διάστημα $(0, 1)$ με το 0 να σημαίνει μηδενική πιθανότητα και το 1 να σημαίνει βεβαιότητα. Η πιθανότητα να φέρει κανείς «3» με ένα ζάρι είναι μία στις έξι, δηλαδή περίπου 0,167 ή 17 τοις εκατό. Στην κατασκευή στατιστικών μοντέλων χρησιμοποιείται αντί της πιθανότητας p μεταξύ άλλων και ο λογάριθμος του λόγου των σχετικών πιθανοτήτων, δηλαδή η πιθανότητα να έρθει «3» δια την πιθανότητα να μην έρθει «3». Ο λόγος αυτός εκφράζεται λογαριθμικά όπου λογάριθμος (\log) είναι ο εκθέτης που πρέπει να υψώσουμε τον λόγο των πιθανοτήτων για να βρούμε έναν συγκεκριμένο αριθμό, την βάση του λογάριθμου. Η έκφραση του λογαρίθμου των λόγων των πιθανοτήτων γράφεται ως $\log\left(\frac{p}{1-p}\right)$ και ονομάστηκε “logit” από τον γιατρό και στατιστικολόγο Joseph



Εικόνα 2
Βαθμός σχετικής δυσκολίας των λέξεων της δοκιμασίας ορθογραφίας (SP1-SP60) όπως υποδεικνύεται από τη βαθμολογία τους στην κλίμακα logit (κάθετος άξονας). Η κατανομή των εξεταζόμενων που έγραψαν σωστά την κάθε λέξη παρουσιάζεται στο αριστερό τμήμα του διαγράμματος (κάθε σύμβολο # αντιστοιχεί σε 14 άτομα)

νότητας του ατόμου (B_p) και της δυσκολίας της ερώτησης (D_i). Λόγου χάρι, αν ένα άτομο που είναι πολύ ικανό (π.χ., σε επίπεδο 2 logit από το μέσο όρο) κληθεί να απαντήσει μια ερώτηση μέ-

τριας δυσκολίας (0 logit), τότε η πιθανότητα να το απαντήσει σωστά θα είναι 0,88. Αν η ερώτηση είναι εύκολη (π.χ., σε επίπεδο -1 logit), τότε η πιθανότητα επιτυχίας για το συγκεκριμένο άτομο

Berkson που έγραψε σχετικά το 1944. Στη logit χρησιμοποιούμε ως βάση των λογάριθμων τον αριθμό $e \approx 2,71828$ και συμβολίζουμε κάποιες φορές με “ln” που σημαίνει «φυσικός λογάριθμος» αντί για “log” που θα μπορούσε να εννοεί μια άλλη βάση. Το μοντέλο Rasch δεν κάνει τίποτα άλλο παρά να μοντελοποιεί τον logit, δηλαδή τις λογιστικές μονάδες. Βασική υπόθεση του μοντέλου Rasch είναι ότι η πιθανότητα ορθής απάντησης σε μία ερώτηση (item) με δύο ενδεχόμενα (σωστό ή λάθος) εξαρτάται από την ικανότητα του μαθητή και τη δυσκολία της ερώτησης με την προϋπόθεση, βέβαια, ότι η απάντηση είναι μόνο θέμα

ικανότητας (ένας παράγοντας) και όχι θέμα άλλων παραγόντων (π.χ. οι συνθήκες διεξαγωγής του τεστ). Οι συγγραφείς παρουσιάζουν στη σελίδα 299 στον μαθηματικό τύπο του μοντέλου Rasch και εξηγούν με παραδείγματα στην επόμενη σελίδα ότι P_{ni} , [η πιθανότητα p του μαθητή n που γράφει τη λέξη i] ($X_{ni} = 1$ [να γράψει σωστά τη συγκεκριμένη λέξη i] / B_n, D_i) [με δεδομένη «/» την ικανότητα B του μαθητή n και τη δυσκολία D της λέξης i] είναι $\frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}}$ [έτσι γίνεται ο τύπος $\log\left(\frac{p}{1-p}\right)$ αν λύσουμε ως προς την πιθανότητα p και λάβουμε υπόψη ότι η πιθανότητα ορθής απάντησης είναι ικανότητα πλην την δυσκολία].

αυξάνεται σε 0,93 (Wright, 1988). Από την εξίσωση (1) προέκυψαν οι δείκτες του βαθμού δυσκολίας κάθε λέξης σε σχέση με τις υπόλοιπες που παρουσιάζονται στην Εικόνα 2. Παρατηρείται ικανοποιητική διασπορά των επιμέρους λέξεων (SP1-SP60) σχεδόν σε ολόκληρο το εύρος της διακύμανσης των τιμών logit (μεταξύ -8 και +8). Οι πρώτες λέξεις έχουν μικρό βαθμό δυσκολίας (καταλαμβάνουν το κάτω μέρος του διαγράμματος), οι λέξεις στη μέση της λίστας παρουσιάζουν μεσαία επίπεδα δυσκολίας, ενώ οι λέξεις στο τέλος της λίστας παρουσιάζουν το μεγαλύτερο βαθμό δυσκολίας για τους εξεταζόμενους, όπως ήταν αναμενόμενο. Εξαιρέση αποτελούν ορισμένες λέξεις όπως η λέξη «ούτε», η οποία, αν και ενδέκατη στη σειρά εμφάνισης στη λίστα κατατάχθηκε τρίτη σε σειρά δυσκολίας. Επομένως και αυτή η ανάλυση συνιστά μια μικρή διαφοροποίηση στη σειρά χορήγησης των λέξεων. Με βάση το βαθμό δυσκολίας που προέκυψε από το μοντέλο Rasch, η προτεινόμενη σειρά για τη δοκιμασία της ορθογραφίας είναι αυτή που παρατίθεται στο Παράρτημα Α.

Στη συνέχεια τα δεδομένα εφαρμόστηκαν σε ένα μοντέλο που ελέγχει την ύπαρξη του σχηματισμού Guttman (όπου η μονάδα υποδεικνύει σωστές απαντήσεις, ενώ το μηδέν λανθασμένες). Στο παράδειγμα του σχηματισμού 1 1 1 1 0 1 0 1 0 0 0 0, ένα άτομο προβλέπεται ότι θα απαντήσει σωστά τις πρώτες τέσσερις ερωτήσεις ενός τεστ. Η προσδοκία για την επίδοση του ατόμου στο μέσο του τεστ (δηλ. σε ερωτήσεις μεσαίας δυσκολίας) είναι ότι κάποιες από αυτές θα τις απαντήσει σωστά, ενώ σε κάποιες άλλες θα κάνει λάθος. Τέλος, η προσδοκία για τις δύσκολες ερωτήσεις ενός τεστ είναι ότι ο «μέσος» εξεταζόμενος δεν θα μπορέσει να τις απαντήσει. Με βάση αυτή την προσδοκία, το μοντέλο Rasch αξιολογεί τόσο τη συμπεριφορά των ατόμων (ανιχνεύοντας αν κάποια άτομα αποκλίνουν συστηματικά από τις προβλέψεις του μοντέλου, π.χ. όταν «κλέβουν» σε ένα τεστ), όσο και των επιμέρους ερωτήσεων ενός τεστ (π.χ., ανιχνεύει αν μια ερώτηση που σχεδιάστηκε για να είναι πολύ δύσκολη έχει τις αναμενόμενες ιδιότητες).

Τα αποτελέσματα από το συνολικό δείγμα παρουσιάζονται στην Εικόνα 3. Όπως φαίνεται στο πάνω τμήμα της εικόνας, οι περισσότερες λέξεις είχαν «στοχαστική» συνεισφορά στη μέτρηση και χαμηλά επίπεδα στατιστικού σφάλματος. Εξαιρέση αποτέλεσαν οι λέξεις 17 και 23, 24, 28, 31, 32, 36, 39, 48, 49, 50, 52, 54, 56 (βλ. Παράρτημα Β για την ταυτότητα αυτών των λέξεων), που παρουσίασαν αποκλίσεις από τις προσδοκίες του μοντέλου Guttman, όπως υποδεικνύεται από υψηλές τιμές του στατιστικού δείκτη *infit*. Δεδομένου ότι αυτός ο στατιστικός δείκτης αξιολογεί τις αποκλίσεις της συμπεριφοράς κάθε λέξης χωριστά σε σχέση με το «μεσαίο» τμήμα του τεστ, αυτό δείχνει ότι οι συγκεκριμένες λέξεις ήταν λιγότερο ή περισσότερο δύσκολες από ό,τι προβλεπόταν με βάση τις προσδοκίες του μοντέλου Guttman σχετικά με τη θέση τους στη λίστα.

Ο δεύτερος δείκτης (*oufit*), που απεικονίζεται στο κάτω μέρος της Εικόνας 3, υποδηλώνει περισσότερο ακραίες αποκλίσεις από το μοντέλο σε σχέση με τις υπόλοιπες λέξεις μόνο για τις λέξεις «φιλί» (Αρ. 17) και «παράδειγμα» (Αρ. 32). Με απλά λόγια, παρόλο που οι εν λόγω λέξεις βρίσκονταν στο μεσαίο τμήμα της λίστας, υπήρχαν αρκετοί μαθητές που τις αντιμετώπισαν παρόμοια με ευκολότερες (προηγούμενες στη λίστα) ή δυσκολότερες λέξεις (επόμενες στη λίστα). Σημειώνουμε πάντως ότι για την απαλοιφή ή τροποποίηση της σειράς χορήγησης των λέξεων απαιτείται εμπειρική επιβεβαίωση της αποκλίνουσας συμπεριφοράς τους σε καινούργιο δείγμα μαθητών.

Διερεύνηση μεροληψίας με το μοντέλο Rasch, δύο παραμέτρων

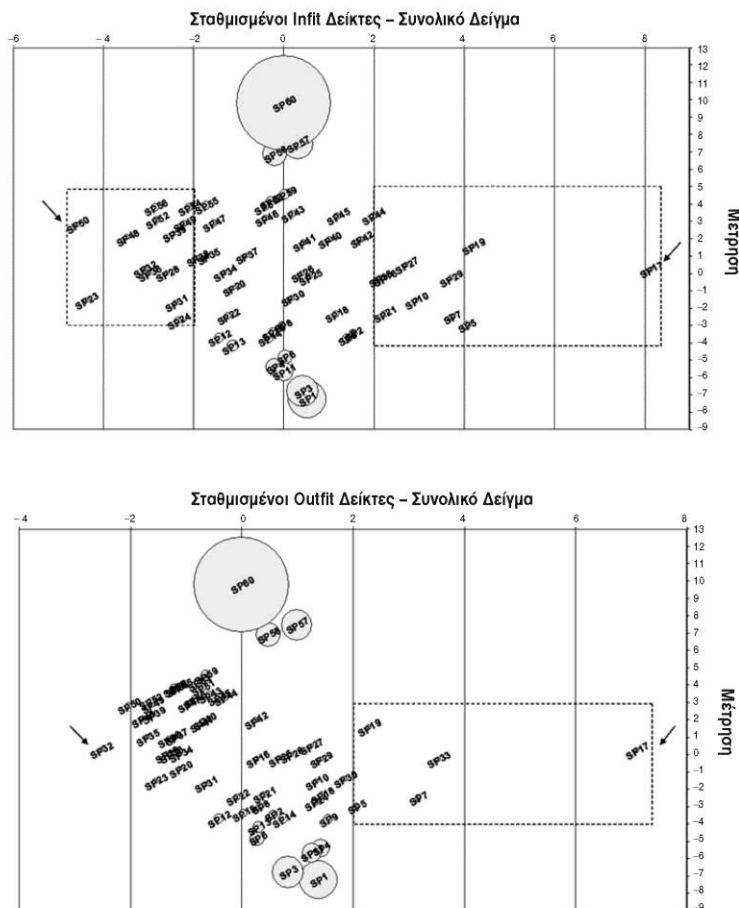
Ο όρος μεροληψία (*bias*) αναφέρεται στην πιθανότητα κάποια λέξη να είναι στατιστικά σημαντικά πιο δύσκολη για κάποιον πληθυσμό σε σχέση με κάποιον άλλο. Ανάλυση Rasch εφαρμόστηκε στα δεδομένα από τη δοκιμασία της ορθογραφίας (πρώτη μέτρηση) για να αξιολογηθεί η ύπαρξη μεροληψίας επιμέρους λέξεων της δοκιμασίας σε σχέση με πολιτισμικούς και γλωσσικούς παράγοντες, συγκρίνοντας το δείγμα μα-

Η Εικόνα 2 στη σελίδα 300 της εργασίας των Σιδερίδη κ.ά. (2008) είναι δείχνει με μονάδες logit τη δυσκολία των ερωτήσεων και ταυτόχρονα το πώς απάντησαν οι μαθητές σε αυτές. Η διακεκομμένη γραμμή στη μέση της Εικόνας 2 στην έρευνα των Σιδερίδη κ.ά. χωρίζει τους μαθητές στ' αριστερά και τις ερωτήσεις στα δεξιά. Οι ερωτήσεις έχουν καταταχθεί κατά σειρά δυσκολίας από τις εύκολες στο κάτω μέρος της Εικόνας 2 στις δύσκολες στο επάνω μέρος της. Ο μέσος όρος δυσκολίας εμφανίζεται με "M" στα δεξιά της γραμμής, ενώ η μία και οι δύο τυπικές αποκλίσεις στη δυσκολία εμφανίζονται με "S" και "T" αντιστοίχως. Κοντά στον μέσο όρο είναι εικοστή έκτη λέξη. Πιο δύσκολη ερώτηση εμφανίζεται να είναι η

εξηκοστή, ενώ φαίνεται στο στην Εικόνα 2 ότι ελάχιστοι μαθητές απαντούν στις πιο δύσκολες λέξεις.

Στη σελίδα 301 οι συγγραφείς ελέγχουν την αρχική κατάταξη των λέξεων από

302 ♦ Γεώργιος Δ. Σιδερίδης, Αγγελική Μουζάκη, Αθανάσιος Πρωτόπαπας & Παναγιώτης Σίμος



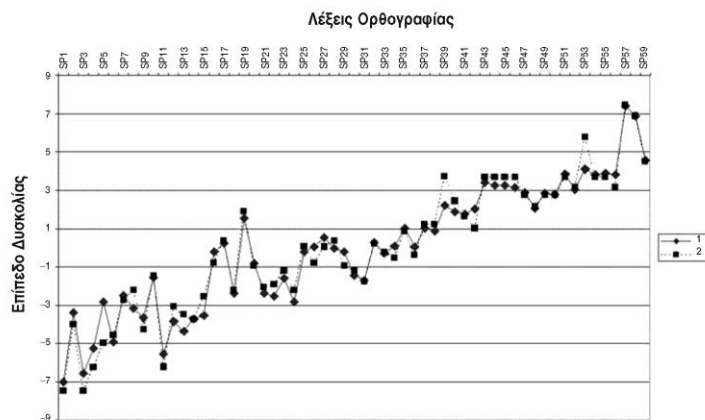
Εικόνα 3

Διαγράμματα «φυσαλίδων» που δείχνουν τη συμπεριφορά καθεμιάς από τις λέξεις της δοκιμασίας, σύμφωνα με το μοντέλο Guttman, στο συνολικό δείγμα (πρώτη μέτρηση). Τα βέλη δείχνουν λέξεις για τις οποίες υπάρχουν υποψίες απρόβλεπτης συμπεριφοράς από την πλευρά των μαθητών, όπως υποδεικνύεται από υψηλές τιμές των στατιστικών δεικτών infit (οριζόντιος άξονας-πάνω εικόνα) και outfit (οριζόντιος άξονας-κάτω εικόνα)

πλευράς δυσκολίας και καταγράφουν αυτές που «παρουσίασαν αποκλίσεις από τις προσδοκίες του μοντέλου Guttman» (ό.π: 301). Στην έννοια του μοντέλου Guttman αναφερθήκαμε κι εμείς στη σελίδα 28 και θυμίζουμε εδώ ότι σύμφωνα με αυτό οι μαθητές απαντούν στις ερωτήσεις που ταιριάζουν με το επίπεδο

ικανότητάς τους και σε όλες τις προηγούμενες (που είναι υποτίθεται ευκολότερες) αλλά όχι στις επόμενες που είναι υποτίθεται δυσκολότερες.

Στην δεύτερη παράγραφο της δεξιάς στήλης της σελίδας 301 οι συγγραφείς (Σιδερίδης κ. ά. 2008) παρουσιάζουν στοιχεία σχετικά με το κατά πόσο το μοντέλο Rasch προσαρμόζεται στα δεδομένα. Αυτό το κάνουν με δύο δείκτες, τον δείκτη Outfit και τον δείκτη Infit, οι οποίοι παρουσιάζονται στη σελίδα 302 της εργασίας των Σιδερίδη κ.ά. (2008) ως σχεδιαγράμματα φυσαλίδων. Τα σχεδιαγράμματα αυτά θα τα εξετάσουμε μόλις δούμε τους δείκτες καλής προσαρμογής του μοντέλου και συγκεκριμένα τους δείκτες **Infit** και **Outfit**. Οι εν λόγω δείκτες καλής προσαρμογής αφορούν τόσο τα ερωτήματα του τεστ (τις 60 λέξεις στη δική μας περίπτωση), όσο και καθέναν από τους συμμετέχοντες μαθητές και μαθήτριες ξεχωριστά. Έτσι, κάθε λέξη και κάθε μαθητής έχουν από τιμή για καθέναν από τους δύο προαναφερθέντες δείκτες.



Εικόνα 4
Βαθμός δυσκολίας κάθε λέξης ανάλογα με τη σειρά παρουσιάσής της στην κλίμακα (SP1-SP60) για το σύνολο του δείγματος (πρώτη μέτρηση). Διακεκομμένη γραμμή: μαθητές με μητρική γλώσσα την ελληνική, συνεχής γραμμή: μαθητές με γονείς μετανάστες. Αντίστοιχα δεδομένα για κάθε τάξη παρουσιάζονται στο Παράρτημα Γ

θητών με ελληνόφωνους γονείς με εκείνο των μεταναστών. Εφαρμόστηκε η διαδικασία ανίχνευσης μεροληψίας DIF (Differential Item Functioning), με την οποία εκτιμήθηκε αν συγκεκριμένες λέξεις ήταν στατιστικώς σημαντικά πιο δύσκολες για τους μαθητές από οικογένειες μεταναστών σε σύγκριση με τους μαθητές που είχαν ως μητρική γλώσσα την ελληνική. Με τη μέθοδο αυτή συγκρίθηκαν οι δύο ομάδες μαθητών στο βαθμό δυσκολίας της κάθε λέξης ξεχωριστά, με παράλληλο έλεγχο της συνολικής επίδοσης στη δοκιμασία. Ως μεροληψία θεωρήθηκε μια στατιστικά σημαντική διαφορά στο βαθμό δυσκολίας μιας λέξης που παρατηρείται σε μέλη διαφορετικών γλωσσικών ομάδων, τα οποία όμως εμφανίζουν την ίδια συνολική βαθμολογία στη δοκιμασία της ορθογραφίας. Για την εκτίμηση πιθανής μεροληψίας χρησιμοποιήθηκε ο δείκτης που

προτάθηκε από τους Wang και Chen (2004) επειδή εκφράζει τη σταθμισμένη (και άρα ουσιαστική) διαφορά μεταξύ των ομάδων που συγκρίνονται. Ο τύπος αυτός έχει ως εξής:

$$t_{11} \frac{d_{i1} - d_{i2}}{\sqrt{s_1^2 + s_2^2}} \quad (2)$$

όπου d_i και d'_i είναι οι πιθανότητες για τους μαθητές κάθε ομάδας να γράψουν σωστά τη λέξη i , και s^2 η διακύμανση κάθε ομάδας. Ο δείκτης αυτός ορίζεται ως η σταθμισμένη μέση διαφορά μεταξύ των ομάδων. Μεγέθη της τάξης του 1,0 θεωρούνται ενδεικτικά μεροληψίας (Wang & Chen, 2004). Όλες οι αναλύσεις πραγματοποιήθηκαν με το λογισμικό Winsteps 3.58 (Linacre, 1999).

Η Εικόνα 4 δείχνει το βαθμό δυσκολίας των λέξεων για κάθε ομάδα ξεχωριστά, όπου φαίνε-

βασιζονται στην ιδέα της κανονικοποιημένης (standardized) διαφοράς μεταξύ της τιμής που παρατηρείται (εδώ η τιμή είναι είτε 0 είτε 1, ανάλογα με το αν ο εξεταζόμενος είχε λάθος ή σωστή την ερώτηση αντιστοίχως) και της τιμής που θα αναμενόταν με βάση το μοντέλο Guttman, για το οποίο μιλήσαμε πιο πάνω. Η αναμενόμενη τιμή υπολογίζεται πιθανολογικά από τα ίδια τα δεδομένα. Το υπόλοιπο αυτής της αφαιρέσεως κανονικοποιείται διαιρούμενο με την τυπική απόκλιση. Τα τετράγωνα των κανονικοποιημένων αυτών διαφορών αθροίζονται είτε για όλους τους εξεταζόμενους είτε για όλα τα ερωτήματα και διαιρούνται με το πλήθος των εξεταζόμενων μαθητών ή των ερωτημάτων, δίνοντας δείκτες

Outfit για εξεταζόμενους ή ερωτήματα αντιστοίχως. Ο τύπος του δείκτη Outfit είναι

$$\frac{1}{N} \sum_{i=1}^N \frac{(U_{ij} - P_{ij})^2}{w_{ij}}, \text{ όπου } U_{ij} \text{ είναι η παρατηρούμενη επίδοση του μαθητή } i \text{ στο}$$

ερώτημα j , P_{ij} η πιθανότητα ορθής απάντησης του μαθητή i στο ερώτημα j , N το μέγεθος του και $w_{ij} = P_{ij}(1 - P_{ij})$ η διακύμανση (de Ayala, 2009).

Στον δείκτη Infit το άθροισμα των τετράγωνων των κανονικοποιημένων

διαφορών διαιρείται με τη συνολική διακύμανση. Είναι δηλαδή
$$\frac{\sum_{i=1}^N (U_{ij} - P_{ij})^2 / w_{ij}}{\sum_{i=1}^N w_{ij}}.$$

Ο δείκτης Outfit επηρεάζεται κυρίως από αποκλίσεις σε ερωτήσεις και εξεταζόμενους που είναι στα άκρα της κλίμακας (εύκολες και δύσκολες ερωτήσεις, μαθητές με υψηλή ή χαμηλή ικανότητα). Αντίθετα, ο δείκτης Infit επηρεάζεται κυρίως από αποκλίσεις στο μεσαίο φάσμα της δυσκολίας των ερωτήσεων ή της ικανότητας των εξεταζόμενων (de Ayala, 2009). Και οι δύο αυτοί δείκτες μπορεί να είναι από μηδέν ως άπειρο και έχουν αναμενόμενη τιμή ίση με τη μονάδα. Οι Σιδερίδης κ.ά. (2008) παρουσιάζουν στη σελίδα 302 της εργασίας τους τις φυσαλίδες για τους δείκτες Outfit και Infit σε σχέση με τα ερωτήματα που χρησιμοποιήθηκαν στην έρευνα. Οι φυσαλίδες αυτές παράγονται από το λογισμικό "Excel" της Microsoft, το οποίο παίρνει τα δεδομένα και τις σχετικές εντολές από το λογισμικό Winsteps. Στους άξονες των τεταγμένων παρουσιάζεται ο βαθμός δυσκολίας των ερωτήσεων σε κλίμακες logit, ενώ στον οριζόντιο άξονα των τετμημένων παρουσιάζονται οι τιμές των δεικτών Infit (επάνω σχήμα) και Outfit (κάτω σχήμα). Οι συγγραφείς (Σιδερίδης, κ. ά. 2008) σημειώνουν με μικρά βέλη τις ερωτήσεις που δεν προσαρμόζονται καλά στο μοντέλο. Αξίζει να σημειωθεί ότι η διάμετρος των φυσαλίδων σχεδιάζεται από το

ται παρόμοιο επίπεδο δυσκολίας για τους δύο πληθυσμούς, εύρημα που υποδεικνύει απουσία μεροληψίας των λέξεων που συνθέτουν τη δοκιμασία. Εξαιρέση αποτέλεσαν οι λέξεις «ξέρω» (αρ. 8) και «ποτιζώ» (αρ. 15) που φάνηκε να είναι πιο εύκολες για τους μαθητές με μητρική γλώσσα την ελληνική, και οι λέξεις «πετάνε» (αρ. 5), «δωρεάν» (αρ. 26) και «οικονομικότερος» (αρ. 42) για τις οποίες βρέθηκε η αντίστροφη τάση. Για τις υπόλοιπες λέξεις δεν παρατηρήθηκε μεροληψία που να υπερβαίνει το επίπεδο στατιστικής σημαντικότητας ($p > 0,05$).

4. Συζήτηση

Στόχος της παρούσας μελέτης ήταν η πληρέστερη ψυχομετρική διερεύνηση των χαρακτηριστικών μιας δοκιμασίας για την αξιολόγηση της ορθογραφικής δεξιότητας κάνοντας χρήση σύγχρονων στατιστικών μεθόδων για τον προσδιορισμό της κατασκευαστικής/δομικής εγκυρότητας, της σχετικής δυσκολίας των επιμέρους λέξεων που την απαρτίζουν, και της ύπαρξης πιθανής μεροληψίας.

Σύμφωνα με τα ευρήματα από την ανάλυση κυρίων συνιστωσών, τα δεδομένα (τουλάχιστον από τις πρώτες τρεις τάξεις) υποστηρίζουν το μοντέλο ενός παράγοντα και επομένως τη μονοδιάστατη δομή της ορθογραφικής δεξιότητας κατά τη γραφή μεμονωμένων λέξεων. Ανάλογο εύρημα έχει αναφερθεί και από σύγχρονη πολύ μεγαλύτερης κλίμακας μελέτη στην ολλανδική ορθογραφία η οποία κάνει χρήση διαφορετικής αναλυτικής προσέγγισης. Συγκεκριμένα οι Notenboom και Reitsma (2003) διερεύνησαν τη δομική εγκυρότητα μιας ορθογραφικής δοκιμασίας (PI-dictee, Geelhoed & Reitsma, 1999), η οποία περιλαμβάνει 135 λέξεις αυξανόμενης ορθογραφικής δυσκολίας και χορηγήθηκε σε 3657 μαθητές του Δημοτικού με αντίστοιχη διαδικασία. Οι ερευνητές χρησιμοποίησαν ανάλυση παραγόντων και μοντέλα IRT (item response theory) τα οποία κατέδειξαν την ύπαρξη ενός παράγοντα που ερμήνευε 40% – 50% της διακύμανσης της

ορθογραφικής δεξιότητας στις τάξεις Β' – Δ', εύρημα που παραπέμπει σε πιθανή εφαρμογή μιας κυρίως στρατηγικής (ορθογραφικής/μορφολογικής) από πολύ νωρίς. Από την άλλη μεριά, η εύρεση ύπαρξης δύο παραγόντων στην Α' Δημοτικού υποδηλώνει την επιπλέον εφαρμογή της φωνολογικής στρατηγικής στα πρώτα στάδια ανάπτυξης της ορθογραφικής δεξιότητας, η οποία όμως φαίνεται να εγκαταλείπεται ήδη από τη Β' τάξη για τους μαθητές της ολλανδικής ορθογραφίας (Notenboom & Reitsma, 2003).

Ένα δεύτερο σημαντικό εύρημα της παρούσας μελέτης ήταν ότι ο βαθμός δυσκολίας των λέξεων ακολουθούσε τις προβλεπόμενες θεωρητικές καταβολές με μικρές εξαιρέσεις, όπως υπέδειξε η εφαρμογή του μοντέλου Rasch. Οι πληροφορίες που αυτό το μοντέλο προσέθεσε στην παρούσα έρευνα είναι πολύ σημαντικές αφού η σειρά των προτάσεων έχει συνέπειες για την αξιολόγηση του μαθητή (εφόσον η κλίμακα χορηγείται με κανόνα διακοπή). Επίσης, η σωστή σειρά των λέξεων μπορεί να έχει συνέπειες στην επιμονή των μαθητών και στην ύπαρξη συναισθημάτων «αβοηθησίας», αν για παράδειγμα αυτοί συναντήσουν λέξεις μεγάλης δυσκολίας στην αρχή του τεστ (Sabatino, 1982. Sideridis, 2003).

Ένα τρίτο σημαντικό εύρημα ήταν ότι σύνθεση της κλίμακας ορθογραφίας δεν φαίνεται να ευνοεί τους Έλληνες μαθητές σε σχέση με τους αλλοδαπούς. Οι αναλύσεις προκατάληψης προτείνουν ότι η κλίμακα είναι έγκυρη και τα αποτελέσματά της μπορούν να χρησιμοποιηθούν και στους δύο πληθυσμούς, εξίσου.

Το ελληνικό εκπαιδευτικό σύστημα παρότι αφιερώνει αξιόλογο διδακτικό χρόνο για την εξοικείωση των μαθητών με τις ορθογραφικές συμβάσεις του γραπτού λόγου δεν παρέχει αντίστοιχα εργαλεία στους εκπαιδευτικούς για πλήρη αξιολόγηση των αναπτυσσόμενων ορθογραφικών δεξιοτήτων με αποτέλεσμα οι εκτιμήσεις τους να βασίζονται είτε σε αναλύσεις ποιοτικών χαρακτηριστικών του παραγόμενου γραπτού είτε σε δοκιμασίες πλήρωσης κριτηρίων (criterion referenced). Κατά αυτό τον τρόπο η ορθογραφία

Excel ανάλογα με τα τυπικά σφάλματα των δεικτών Infit και Outfit της κάθε ερώτησης. Όπως θα εξηγήσουμε, οι εν λόγω δείκτες δεν ακολουθούν στις φυσαλίδες την κατανομή χι τετράγωνο, αλλά την κατανομή t για τον έλεγχο της στατιστικής τους σημαντικότητας.

Συνήθως οι τιμές outfit και infit παρουσιάζονται σε δύο μορφές στα στατιστικά πακέτα. Είτε ως mean square statistics (MNSQ) είτε ως standardized statistics (ZSTD) (Wright & Pancharakesan, 1969). Όπως είδαμε, ο δείκτης Outfit βασίζεται στο άθροισμα των τετραγώνων των τυποποιημένων καταλοίπων. Η κατανομή των τυποποιημένων υπολοίπων του μοντέλου Rasch είναι κατά προσέγγιση κανονική. Το άθροισμα των τετραγώνων των υπολοίπων ακολουθεί την κατανομή χι τετράγωνο. Η διαίρεση αυτού του αθροίσματος με

τους βαθμούς ελευθερίας δίνει το μέσο όρο αυτού του αθροίσματος, δηλαδή το OUTFIT MEANSQ, με μέσο όρο 1 και εύρος από 0 μέχρι άπειρο. Τιμές μεγαλύτερες του 1 δείχνουν «θόρυβο» στο μοντέλο Rasch. Οι τιμές OUTFIT MEANSQ είναι φυσικά σε κλίμακα ίσων λόγων κι έτσι μια τιμή π.χ. MEANSQ 1,2 σημαίνει 20% περισσότερο «θόρυβο». Τιμές μικρότερες του 1 καταδεικνύουν «υπερπροσαρμογή» (overfit) των δεδομένων στο μοντέλο Rasch, πράγμα που σημαίνει ότι οι αποκρίσεις των συμμετεχόντων είναι πολύ πιο προβλέψιμες ή τακτοποιημένες από ό,τι θα έπρεπε. Ο δείκτης Infit είναι ένας Outfit σταθμισμένος με βάση την «πληροφορία» που προσφέρει κάθε ερώτηση. Η στάθμιση έρχεται να μειώσει την επίδραση όσων αποκρίσεων συνδέονται με μικρή διακύμανση ή είναι εκτός στόχου και σε γενικές γραμμές δεν δίνουν πολύ πληροφορία στο μοντέλο. Οι σχεδιαστές του λογισμικού Winsteps χρησιμοποιούν μια εναλλακτική μορφή αυτών των δύο δεικτών, η οποία προτάθηκε από τους Wilson και Hilferty (1931).

DIFFERENTIAL ITEM FUNCTIONING

Στη συνέχεια, οι συγγραφείς προχωρούν στον έλεγχο της πιθανής διαφοροποιητικής λειτουργίας των ερωτήσεων του τεστ (differential item functioning). Στον έλεγχο δηλαδή του αν τα ερωτήματα είναι το ίδιο δύσκολα για διαφορετικές ομάδες εξεταζόμενων. Σύμφωνα με τους Clauser και Mazor (1998), διαφοροποιητική λειτουργία έχουμε όταν δύο εξεταζόμενοι στο ίδιο επίπεδο ικανότητας αλλά από διαφορετικά γκρουπ έχουν διαφορετικές πιθανότητες να απαντήσουν σωστά σε μια ερώτηση. Αν ο παράγοντας που προκαλεί αυτή τη διαφοροποίηση ανάμεσα στα δύο γκρουπ δεν είναι μέρος της θεωρητικής κατασκευής που ελέγχεται από το τεστ τότε το τεστ είναι μεροληπτικό. Από μόνη της δηλαδή μια διαφοροποίηση ανάμεσα σε εξεταζόμενους δεν σημαίνει ότι το τεστ είναι κατ' ανάγκη και μεροληπτικό. Για να θεωρηθεί το τεστ μεροληπτικό θα πρέπει η διαφοροποιητική λειτουργία να εμφανίζεται σε τομέα που δεν συνδέεται με την θεωρητική κατασκευή που ελέγχεται. Αν η διαφοροποιητική λειτουργία εμφανίζεται σε τομείς που σχετίζονται με την προς έλεγχο θεωρητική κατασκευή τότε δεν μιλάμε για «μεροληψία» του τεστ αλλά για «επίδραση» της προς έλεγχο μεταβλητής.

Η απόφαση για το αν έχουμε μεροληψία ή επίδραση δεν είναι πάντα εύκολη και συνήθως απαιτείται ομάδα ειδικών περί της προς έρευνα θεωρητικής κατασκευής για να αποφανθούν. Οι δύο ομάδες στη συγκεκριμένη περίπτωση ήταν οι μαθητές που έχουν γονείς μετανάστες και οι μαθητές που δεν έχουν γονείς μετανάστες και η θεωρητική κατασκευή είναι η γλωσσική ικανότητα. Πράγματι, το γλωσσικό

κή επίδοση των μαθητών δεν επιδέχεται σύγκριση, με κάποιο σχετικό δείγμα αναφοράς, η οποία είναι συχνά απαραίτητη για την εξαγωγή συμπερασμάτων που αναφέρονται στην κατάκτηση στόχων του αναλυτικού προγράμματος ή στόχων που έχουν τεθεί σε πλαίσιο εξατομικευμένων προγραμμάτων παρέμβασης για μαθητές με διαταχές του γραπτού λόγου. Η σημασία της ορθογραφικής δεξιότητας στη γενικότερη ακαδημαϊκή επίδοση έγκειται μεταξύ άλλων στο γεγονός ότι δημιουργεί γέφυρες μεταξύ εσωτερικής (implicit, άρρητης) και εκφραστικής (explicit, ρητής) φωνολογικής ενημερότητας βοηθώντας τους μαθητές να κατανοήσουν τη δομή της ομιλούμενης γλώσσας (Cataldo & Ellis, 1988). Σύμφωνα με τους Cataldo και Ellis τα είδη αυτά της φωνολογικής ενημερότητας καθορίζουν τα επίπεδα αναγνωστικής ικανότητας, μέσα από την ορθογραφική ικανότητα. Σε μια σειρά από μελέτες η Gregg και οι συνεργάτες της βρήκαν ότι τα μεγαλύτερα προβλήματα γλώσσας μετά το Λύκειο εντοπίζονται στην ορθογραφία, στην αποκωδικοποίηση και στην ευχέρεια (Gregg et al., 2005, Gregg et al., 2002). Επομένως η απόκτηση δεξιοτήτων ορθογραφίας είναι εξαιρετικά σημαντική για την ανάπτυξη άλλων δεξιοτήτων και για την εκτίμηση αυτών χρειάζονται ψυχομετρικά έγκυρα εργαλεία. Η χρήση έγκυρων και αξιόπιστων εργαλείων μέτρησης που συνοδεύονται από δεδομένα στάθμισης είναι αναγκαία και επιβεβλημένη στο σύγχρονο σχολείο σαν τμήμα μιας αμερόληπτης αξιολογικής διαδικασίας η οποία προσφέρει αναγκαία πληροφόρηση για το σχεδιασμό διδασκαλίας υψηλής αποτελεσματικότητας. Η υπό μελέτη κλίμακα προσφέρεται για τη συγκεκριμένη χρήση καθώς φαίνεται να πληροί σημαντικά ψυχομετρικά χαρακτηριστικά (Μουζάκη κ.ά., 2007) τα οποία, όπως έχει προαναφερθεί, αφορούν (i) υψηλή εσωτερική συνέπεια, (ii) υψηλή αξιοπιστία επανεξέτασης, (iii) μονοδιάστατη σύνθεση, (iv) ικανοποιητική διάταξη των επιμέρους λέξεων κατά σειρά δυσκολίας και (v) χαμηλά επίπεδα μεροληψίας στη χρήση του σε ομάδες του πληθυσμού με διαφορετικό γλωσσικό υπόβαθρο.

Πέραν όμως της δυνατότητας εξαγωγής ποσοτικού τύπου συμπερασμάτων σχετικά με το βαθμό ανάπτυξης της ορθογραφικής δεξιότητας, η συγκεκριμένη κλίμακα προσφέρεται και για ποιοτικές αναλύσεις οι οποίες είναι ιδιαίτερες χρήσιμες για το σχεδιασμό του περιεχομένου διδακτικών παρεμβάσεων στην ορθογραφία (Moats, 1995, Treiman, 1993). Οι αναλύσεις της λανθασμένης απόδοσης των λέξεων και της συχνότητας εμφάνισης συγκεκριμένων λαθών συμβάλλουν στη διαμόρφωση ενός πλήρους προγράμματος μαθητικών αδυναμιών και δυνατοτήτων που παραπέμπει συχνά σε συγκεκριμένο αναπτυξιακό ορθογραφικό στάδιο και τη χρήση αντίστοιχων στρατηγικών παραγωγής γραπτού λόγου που συνδέονται με αυτό (Bear et al., 1996). Αυτό συμβαίνει γιατί η εν λόγω δοκιμασία περιλαμβάνει όχι μόνο μια ευρεία σύνθεση ορθογραφικών μοτίβων συχνών στην ελληνική γλώσσα, αλλά και λέξεις που αποτελούν εξαίρεση σε γνωστούς κανόνες επιχειρώντας μια ισόρροπη αντιπροσώπευση της ποικιλίας των λεκτικών μορφημάτων που παρουσιάζονται στην ελληνική ορθογραφία. Η διάταξη των υπό εξέταση λέξεων όχι μόνο ανάλογα με το μήκος τους (μονοσύλλαβες προς πολυσύλλαβες), αλλά και σύμφωνα με το βαθμό δυσκολίας τους σε συνδυασμό με τον κανόνα διακοπής (στις περιπτώσεις διαδοχικών λαθών) οριοθετεί την εξέταση κατά τρόπο ευνοϊκό προς το μαθητή αλλά ταυτόχρονα και αποτελεσματικό ως προς το χρόνο χορήγησης.

Η παρούσα μελέτη έχει επίσης κάποιους περιορισμούς. Αν και το συμπέρασμα που προέκυψε ήταν ότι το συγκεκριμένο τεστ ορθογραφίας είναι έγκυρο, εντούτοις πρέπει να σημειωθεί αφενός ότι δεν έχει ελεγχθεί η τρέχουσα και προβλεπτική εγκυρότητα του εν λόγω εργαλείου, γεγονός που θα ενίσχυε την κλινική χρησιμότητά του, αφετέρου ότι η εγκυρότητα δεν αποτελεί στατικό των τεστ αλλά των δειγμάτων. Δηλαδή, το γεγονός ότι τα ψυχομετρικά χαρακτηριστικά αυτής της κλίμακας ήταν ικανοποιητικά δεν σημαίνει ότι θα συμβεί το ίδιο και με μεταγενέστερα δείγματα. Αν και η προσδοκία είναι ότι, εφόσον τα

background των εξεταζομένων είναι παράγοντας με σημαντική επίπτωση στη μελέτη της γλωσσικής ικανότητάς τους βλ. σχ. (Chen & Henning, 1985; Brown, 1999; Elder, 1996; Kim 2001. Ο έλεγχος αυτός είναι πολύ σημαντικός ιδιαίτερα σε μια γλωσσική δοκιμασία.

Η διαφοροποίηση για την οποία μιλάμε υπολογίζεται από ένα t-test, το οποίο προτάθηκε από τους Wright and Stone (1979). Ο τύπος του συγκεκριμένου t-test δίνεται από τους συγγραφείς και είναι: $t = \frac{d_{i_2} - d_{i_1}}{\sqrt{s_{2i_2} - s_{2i_1}}}$ όπου d_{i_1} είναι η δυσκολία του ερωτήματος i στην μία ομάδα 1, d_{i_2} είναι η δυσκολία του ερωτήματος

i στην άλλη ομάδα, s_{2i_1} το τυπικό σφάλμα του d_{i_1} και s_{2i_2} το τυπικό σφάλμα του d_{i_2} . Με βάση αυτό το τεστ οι ερευνητές δεν βρήκαν στατιστικώς σημαντική διαφορά ανάμεσα σε μαθητές που είχαν ως μητρική γλώσσα την ελληνική και μαθητές που δεν είχαν ως μητρική γλώσσα την ελληνική. Μάλιστα, οι λέξεις «πετάνε», «δωρεάν» και «οικονομικότερος» βρέθηκε να είναι πιο εύκολες για μαθητές που δεν είχαν την ελληνική ως μητρική γλώσσα.

Στην ενότητα της «συζήτησης», οι συγγραφείς αναφέρουν ξανά τους αρχικούς

306 ♦ Γεώργιος Δ. Σιδερίδης, Αγγελική Μουζάκη, Αθανάσιος Πρωτόπαπας & Παναγιώτης Σίμος

μεταγενέστερα δείγματα αποτελούν τυχαία δείγματα από τον ίδιο πληθυσμό, τα ψυχομετρικά χαρακτηριστικά του συγκεκριμένου τεστ θα είναι παρόμοια, εντούτοις συνιστάται επιφύλαξη στην εξαγωγή συμπερασμάτων.

Βιβλιογραφία

- Αϊδίνης, Α. (2006). Η ανάπτυξη της φωνολογικής ενημερότητας και η σχέση της με την απόκτηση του γραμματισμού. *Επιστημονική Επιτηρίδα της Ψυχολογικής Εταιρείας Βορείου Ελλάδος*, 4, 17-42.
- Αϊδίνης, Α. & Παράσχου, Δ. (2004). Ανάπτυξη της ορθογραφημένης γραφής στο δημοτικό σχολείο: Λάθη και στρατηγικές των παιδιών. Στο Ν. Μακρής & Δ. Δεσλή (Επιμ.). *Η γνωστική ψυχολογία σήμερα: Γέφυρες για τη μελέτη της νόησης* (σσ. 245-252). Αθήνα: Τυπωθήτω-Δαρδανός.
- Bakamidis, S. & Carayannis, G. (1987). "Phonemia": A phoneme transcription system for speech synthesis in Modern Greek. *Speech Communication*, 6, 159-169.
- Bear, D. R., Invernizzi, M., Templeton, S. & Johnston, F. (1996). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Upper Saddle River, NJ: Merrill.
- Bentler, P. M. (2004). *EQS Structural equations program manual*. Los Angeles, CA: Scientific Software International.
- Bentler, P. M. (in press). *EQS 6 Structural equations program manual*. Los Angeles, CA: Scientific Software International.
- Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skill to kindergarten and first-grade reading achievement. *Journal of Educational Psychology*, 76, 610-622.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Bradley, L. & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, 30, 419-421.
- Caravolas, M., Hulme, C. & Snowling, M. (2001). The foundations of spelling ability: Evidence from a three-year longitudinal study. *Journal of Memory and Language*, 45, 751-774.
- Cataldo, S. & Ellis, N. (1988). Interactions in the development of spelling and phonological skills. *Journal of Research in Reading*, 11, 2, 86-109.
- Chliounaki, K. & Bryant, P. (2002). Construction and learning to spell. *Cognitive Development*, 17, 1489-1499.
- Cossu, G., Gugliotta, M. & Marshall, J. C. (1995). Acquisition of reading and written spelling in a transparent orthography: Two non parallel processes? *Reading & Writing: An Interdisciplinary Journal*, 7, 9-22.
- Γεώργας, Δ., Παρασκευόπουλος, Ι. Ν., Μπεζεβέγκης, Ι. Γ. & Γιαννίτσας, Ν. Δ. (1997). *Ελληνικό WISC-III: Wechsler Κλίμακες Νοημοσύνης για Παιδιά*. Αθήνα: Ελληνικά Γράμματα.
- Geelhoed, J. & Reitsma, P. (1999). *PI-dictee*. Swets & Zeitlinger, Lisse, Switzerland.
- Gorsuch, T. (1983). *Factor analysis*. New York: Erlbaum.
- Gregg, N., Hoy, C., Flaherty, D., Norris, P., Coleman, C., Davis, M. & Jordan, M. (2005). Documenting decoding and spelling accommodations for post-secondary students demonstrating dyslexia: It's more than processing speed. *Learning Disabilities: A Contemporary Journal*, 3, 1-17.
- Gregg, N., Coleman, C., Davis, M., Stennet, R., Nielson, K., Knight, D. & Hoy, C. (2002). Lexical and sublexical processing of adults with learning disorders and attention deficit/hyperactivity disorders. In E. Witruck & T. Lachbach (Eds), *Basic functions of language and language disorders* (pp. 31-60). Netherlands: Kluwer Academic Publisher.
- Διακογιώργη, Κ., Μπαρής, Θ. & Βαλμάς, Θ. (2005). Ικανότητα χρήσης μορφολογικών στρατηγικών στην ορθογραφημένη γραφή από μαθητές της Α' τάξης του δημοτικού. *Ψυχολογία*, 12, 568-586.
- Ehri, L. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167-188.
- Holmes, V. M. & Babauta, M. L. (2005). Single or dual representations for reading and spelling? *Reading and Writing*, 18, 257-280.

στόχους της έρευνας, ώστε να μπορέσουν να τους συνδέσουν με τα συμπεράσματα της μελέτης. Τα τρία κυριότερα συμπεράσματα παρουσιάζονται σε τρεις διαφορετικές παραγράφους στις οποίες γίνεται παρουσίαση νέων πληροφοριών από τη διεθνή βιβλιογραφία με μοναδικό σκοπό, όμως, όχι να πληροφορήσει τον αναγνώστη για την υπάρχουσα βιβλιογραφία σχετικά με το

θέμα, αλλά για να συνδέσει τα ευρήματα με άλλα παρόμοια ευρήματα ερευνών στον ευρωπαϊκό χώρο. Οι συγγραφείς αναφέρουν επίσης της επιστημολογικές τους θέσεις γράφοντας για χρήση «έγκυρων και αξιόπιστων εργαλείων μέτρησης» καθώς και «δεδομένα στάθμισης» τα οποία είναι «επιβεβλημένα στο σύγχρονο σχολείο σαν τμήμα μιας αμερόληπτης αξιολογικής διαδικασίας η οποία προσφέρει αναγκαία πληροφόρηση για τον σχεδιασμό διδασκαλίας υψηλής αποτελεσματικότητας». Στην τελευταία παράγραφο οι συγγραφείς αναφέρονται στους «περιορισμούς» της έρευνας αναφερόμενοι στην «τρέχουσα και προβλεπτική εγκυρότητα του εν λόγω εργαλείου». «Τρέχουσα εγκυρότητα» ενός τεστ είναι ο βαθμός συσχέτισης των απαντήσεων σε

αυτό με τις απαντήσεις των ίδιων εξεταζόμενων σε ένα άλλο τεστ που τεκμηριωμένα και αναγνωρισμένα μετράει την ίδια θεωρητική κατασκευή, ενώ «προβλεπτική» εγκυρότητα είναι ο βαθμός συσχέτισης των αποτελεσμάτων ενός τεστ με τα αποτελέσματα ενός άλλου τεστ που γίνεται στο μέλλον και το οποίο μετράει με αδιαμφισβήτητο τρόπο την θεωρητική κατασκευή. Τρέχουσα εγκυρότητα έχει ένα τεστ γνώσεων της Αγγλικής γλώσσας το οποίο ταιριάζει με

ένα τεστ που αποδεδειγμένα μετράει καλά τη γνώση της Αγγλικής, ενώ προβλεπτική εγκυρότητα έχει όταν προβλέψει σωστά ποιοι μαθητές θα πάρουν

το

- Joreskog, K. G. & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.
- Juel, C., Griffith, P. & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78, 243-255.
- Καρατζάς, Α. Δ. (2005). *Μάθηση της ορθογραφικής δεξιότητας. Γνωστικο-ψυχολογιστική προσέγγιση*. Αθήνα: Γρηγόρης.
- Linacre, J. M. (1999). *A user's guide and manual to winsteps*. Chicago: MESA press.
- Linacre, J. M. & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, 8, 360-361.
- Ludlow, L. H. & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement*, 55, 967-975.
- Lundberg, I., Olofsson, A. & Wall, S. (1980). Reading and spelling skill in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology*, 21, 159-173.
- Moats, L. C. (2005). How spelling supports reading. And why it is more regular and predictable than you may think. *American Educator*, 6, 12-43.
- Moats, L. C. (1995). *Spelling: Development, disability and instruction*. Baltimore, MD: York Press.
- Moustaki, I. (2001). A review of exploratory factor analysis for ordinal categorical data. In R. Cudeck, S. du Toit & D. Sorbom (Eds), *Structural equation modelling: Present and future* (pp. 461-480). Lincolnwood, IL: Scientific Software International.
- Mouzaki, A. & Sideridis, G. (2007). Profiles of poor readers among Greek students of elementary school. *Hellenic Journal of Psychology*, 4, 205-232.
- Μουζάκη, Α., Πρωτόπαπας, Α., Σιδερίδης, Γ. Δ. & Σίμος, Π. (2007). Διερεύνηση των ψυχομετρικών χαρακτηριστικών μιας δοκιμασίας ορθογραφικής δεξιότητας μαθητών της Β, Γ, Δ, και Ε τάξης του δημοτικού σχολείου. *Επιστήμες της Αγωγής*, 1, 129-146.
- Nation, K. & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly*, 32, 154-167.
- Nikolopoulos, D., Goulandris, N. & Snowling, M. (2003). Developmental dyslexia in Greek. In N. Goulandris (Ed.), *Dyslexia in different languages: Cross-linguistic comparison* (pp. 53-57). London: Whurr.
- Notenboom, A. & Reitsma, P. (2003). Investigating the dimensions of spelling ability. *Educational & Psychological Measurement*, 63, 1039-1059.
- Nunes, T., Adinís, A. & Bryant, P. (2006). The acquisition of written morphology in Greek. In R. M. Joshi & P. G. Aaron (Eds), *Handbook of orthography and literacy* (pp. 201-218). Mahwah, New Jersey: Lawrence Erlbaum.
- Porpodas, C. D. (2001). Cognitive processes in first grade reading and spelling of Greek. *Ψυχολογία*, 8, 384-400.
- Porpodas, C. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities*, 32, 406-417.
- Πόρποδας, Κ. (1992). Η εκμάθηση της ανάγνωσης και της ορθογραφίας σε σχέση με την ηλικία και τη φωνημική επίγνωση. *Ψυχολογία*, 1, 30-43.
- Protopapas, A. & Skaloumbakas, C. (2007). Traditional and computer based screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities*, 40, 15-36.
- Πρωτόπαπας, Α. & Σκαλούμπακας, Χ. (2008). Η αξιολόγηση της αναγνωστικής ευχέρειας για τον εντοπισμό αναγνωστικών δυσκολιών. *Ψυχολογία*, 15, 267-289.
- Protopapas, A., Sideridis, G. D., Simos, P. G. & Mouzaki, A. (2007). The development of lexical mediation in the relationship between text comprehension and word reading skills in Greek. *Scientific Studies of Reading*, 11, 165-197.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Sabatino, D. A. (1982). Research on achievement motivation with learning disabled populations. *Ad-*

Certificate of Proficiency in English. Στην παρουσίαση της βιβλιογραφίας οι συγγραφείς ακολουθούν το στυλ της Αμερικάνικης Ψυχολογικής Εταιρίας (American Psychological Association), αλλά χωρίς να αναφέρουν τη διεύθυνση doi κάθε άρθρου. Η διεύθυνση αυτή είναι η μόνιμη θέση ενός ηλεκτρονικού κειμένου μέσα στο Διαδίκτυο (digital object identifier), ώστε ένας πιθανός μελετητής να μπορεί να εντοπίσει το άρθρο.

- vances in Learning and Behavioral Disabilities 1, 75-116.
- Seymour, P. H. K, Aro, M. & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *The British Journal of Psychology*, 94, 143-174.
- Sideridis, G. D. (2003). On the origins of helpless behavior in students with learning disabilities: Avoidance motivation? *International Journal of Educational Research*, 39, 497-517.
- Sideridis, G., Mouzaki, A., Simos, P. & Protopapas, A. (2006). Classification of students with reading comprehension difficulties: The roles of motivation, affect, and psychopathology. *Learning Disability Quarterly*, 29, 159-180.
- Smith, E. V. & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models and applications*. Maple Grove, MN: JAM Press.
- Snow, C., Griffin, P. & Burns, M. S. (2005). *Knowledge to Support the Teaching of Reading: Preparing Teachers for a Changing World*. San Francisco: Jossey-Bass.
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.
- Wang, W. & Chen, H. (2004). The standardized mean difference within the framework of item response theory. *Educational and Psychological Measurement*, 64, 201-223.
- Wright, B. D. (1993). "Logits". *Rasch Measurement Transactions*, 7, 288.
- Wright, B. D. (1988). Some comments about guessing. *Rasch Measurement Transactions*, 1, 9.
- Wright, B. D. & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.

ΠΑΡΑΡΤΗΜΑΤΑ

Ψυχομετρική διερεύνηση μιας ορθογραφικής δοκιμασίας ♦ 309

Παράρτημα Α

Προτεινόμενη σειρά λέξεων ορθογραφίας που προέκυψε από την ανάλυση δυσκολίας των λέξεων με το μοντέλο Rasch για το συνολικό δείγμα.

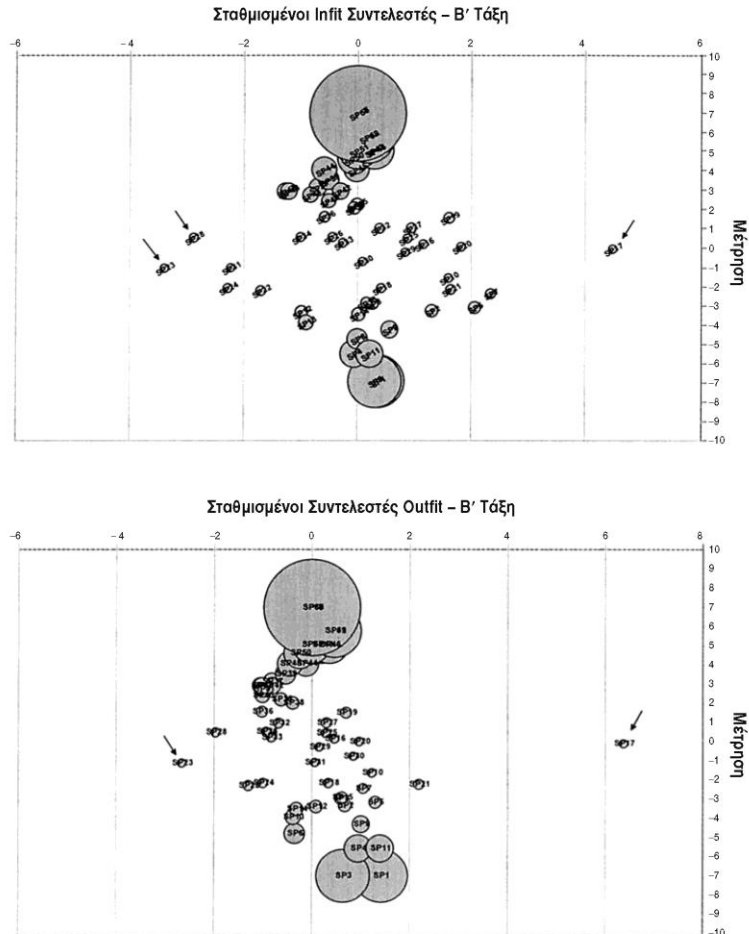
Σειρά λέξης			Συντελεστής Δυσκολίας*	Σειρά λέξης			Συντελεστής Δυσκολίας*
A/A	Αρχική	Τελική		A/A	Αρχική	Τελική	
1	από	από	-7,24	31	αυτοκίνητο	άδεια	0,04
2	έλα	και	-6,79	32	παράδειγμα	φιλί	0,24
3	και	ούτε	-5,67	33	φωτισμένος	παράδειγμα	0,26
4	είναι	είναι	-5,41	34	άδεια	παχτών	0,48
5	πετάνε	άλογο	-4,84	35	ζεσταίνει	δανείζω	0,89
6	άλογο	αυτός	-4,17	36	ονειρεύομαι	ζεσταίνει	1,00
7	φιλος	έτσι	-3,75	37	αλλιώτικος	αλλιώτικος	1,03
8	ξέρω	τραπέζι	-3,71	38	δανείζω	δίχτυ	1,56
9	έτσι	λέω	-3,71	39	μόλυνση	πετρέλαιο	1,76
10	όμως	έλα	-3,47	40	πρακτορείο	πρακτορείο	1,92
11	ούτε	ποτιζώ	-3,38	41	πετρέλαιο	οικονομικότερος	1,96
12	λέω	ξέρω	-3,01	42	οικονομικότερος	ηθοποιός	2,06
13	αυτός	πετάνε	-3,01	43	υποχρεωμένος	μόλυνση	2,27
14	τραπέζι	μεγαλώνω	-2,73	44	αναχώρηση	αντικείμενο	2,79
15	ποτιζώ	φιλος	-2,52	45	χαρακτηριστικό	διευθυντής	2,85
16	πηγή	ζωγραφική	-2,46	46	ξεφυλλίζοντας	κυνηγητό	2,88
17	φιλί	φυτό	-2,34	47	κυνηγητό	εκατομμύριο	3,06
18	φυτό	πέτρινο	-2,33	48	ηθοποιός	ξεφυλλίζοντας	3,16
19	δίχτυ	αυτοκίνητο	-1,70	49	διευθυντής	χαρακτηριστικό	3,28
20	χρήματα	πηγαίνω	-1,55	50	αντικείμενο	αναχώρηση	3,28
21	πέτρινο	όμως	-1,55	51	επιχείρηση	υποχρεωμένος	3,42
22	ζωγραφική	επιβάτης	-1,41	52	εκατομμύριο	νοημοσύνη	3,79
23	πηγαίνω	χρήματα	-0,82	53	ειδοποιήθηκε	αισθήσεις	3,79
24	μεγαλώνω	χώρισα	-0,28	54	αισθήσεις	επιχείρηση	3,85
25	σοφοί	φωτισμένος	-0,27	55	νοημοσύνη	αποχαιρέτησα	3,90
26	δωρεάν	πηγή	-0,25	56	αποχαιρέτησα	ειδοποιήθηκε	4,15
27	παχτών	σοφοί	-0,20	57	χείμαρρος	ματαιώνεται	4,54
28	ταμείο	δωρεάν	-0,02	58	εγχειριστή	εγχειριστή	6,88
29	χώρισα	ταμείο	0,01	59	ματαιώνεται	χείμαρρος	7,44
30	επιβάτης	ονειρεύομαι	0,02	60	αποδεικνύω	αποδεικνύω	9,78

* Κλίμακα logit.

Και εργασία των Σιδερίδη κ.ά. (2008) κλείνει με τρία παραρτήματα. Στο πρώτο, σελίδα 307 στο πρωτότυπο κείμενο, οι συγγραφείς παρουσιάζουν τον τελικό κατάλογο των λέξεων τοποθετημένες σε σειρά δυσκολίας σύμφωνα με το μοντέλο Rasch, συνοδευόμενες από τον βαθμό δυσκολία τους, πάντα κατά ανάλυση Rasch.

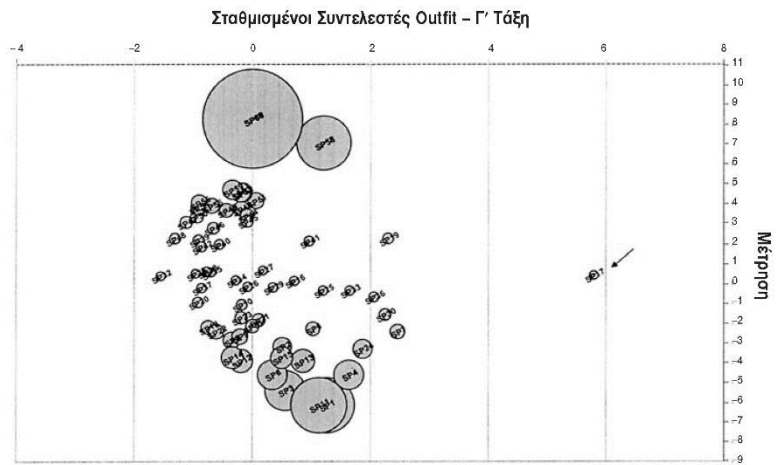
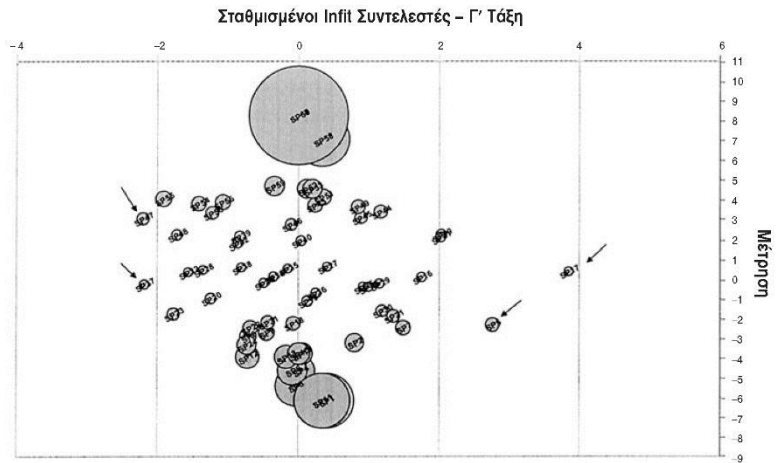
Παράρτημα Β

Διαγράμματα «φυσαλίδων» που δείχνουν τη συμφωνία των λέξεων με το μοντέλο Guttman ανά τάξη (πρώτη μέτρηση). Τα βέλη δείχνουν λέξεις για τις οποίες υπάρχουν υποψίες απρόβλεπτης συμπεριφοράς από την πλευρά των μαθητών, όπως υποδεικνύεται από υψηλές τιμές των στατιστικών δεικτών infit (οριζόντιος άξονας-πάνω εικόνα κάθε ζευγους) και outfit (οριζόντιος άξονας-κάτω εικόνα).

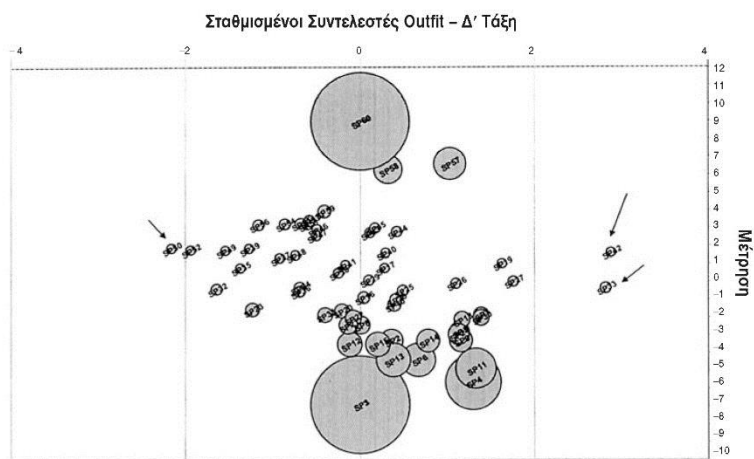
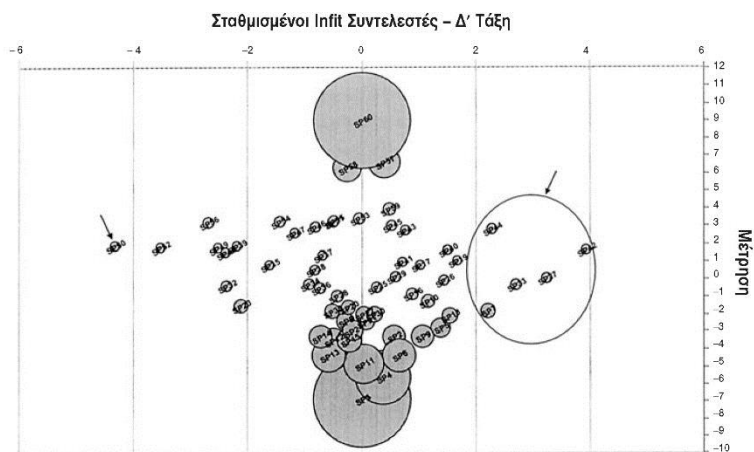


Στο Παράρτημα Β, οι συγγραφείς παρουσιάζουν τα διαγράμματα φυσαλίδων. Όπως έχουμε αναφέρει, η διάμετρος τους είναι ανάλογη των τυπικών σφαλμάτων των κανονικοποιημένων δεικτών Infit και Outfit της. Τέλος, στο Παράρτημα Γ, οι συγγραφείς παρουσιάζουν σε λογιστικές μονάδες το επίπεδο δυσκολίας κάθε λέξης σε κάθε τάξη που έγινε η έρευνα, για μαθητές που είχαν και για μαθητές που δεν είχαν τη μητρική ως ξένη γλώσσα. Η εργασία κλείνει με την περίληψη στην αγγλική γλώσσα.

Παράρτημα Β
(συνέχεια)



Παράρτημα Β
(συνέχεια)

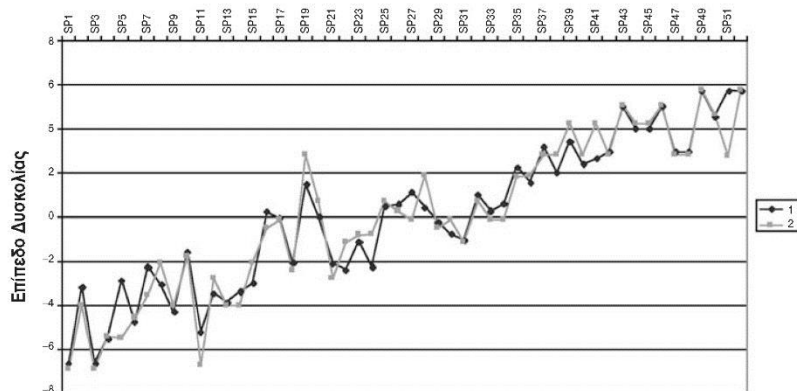


Παράρτημα Γ

Βαθμός δυσκολίας κάθε λέξης ανάλογα με τη σειρά παρουσιάσής της στην κλίμακα (SP1–SP60) για κάθε τάξη (πρώτη μέτρηση). (1): μαθητές με μητρική γλώσσα την ελληνική, (2): μαθητές με γονείς μετανάστες.

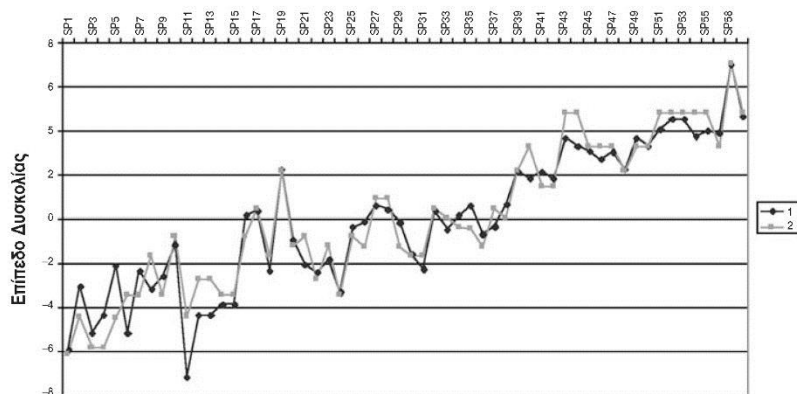
Εικόνα DIF – Β' Τάξη

Λέξεις Ορθογραφίας



Εικόνα DIF – Γ' Τάξη

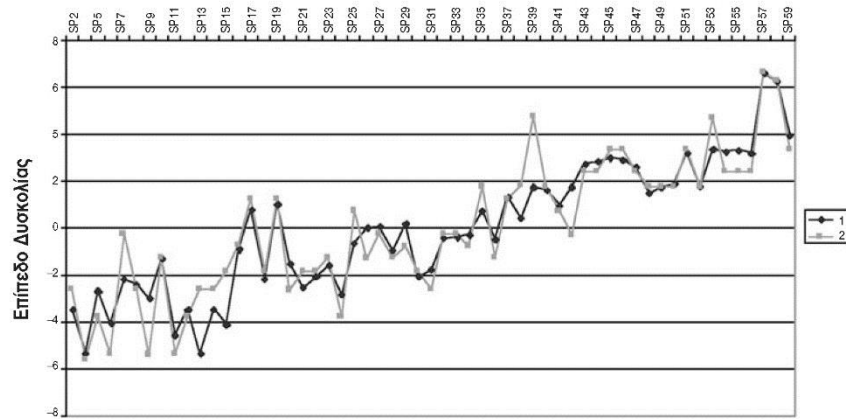
Λέξεις Ορθογραφίας



Παράρτημα Γ
(συνέχεια)

Εικόνα DIF – Δ' Τάξη

Λέξεις Ορθογραφίας



Psychometric evaluation of a spelling test for elementary school students

GEORGIOS D. SIDERIDIS¹, ANGELIKI MOUZAKI,
ATHANASSIOS PROTOPAPAS & PANAGIOTIS SIMOS

ABSTRACT

The purpose of the present study was to examine the psychometric properties of a new measure of spelling achievement. Using a large sample of schoolchildren from Grades 2–4 (N=580), results pointed to the existence of one latent spelling dimension, which was verified across grade and using multiple measurement points. Furthermore, indices of reliability and internal consistency were adequate suggesting the stability of the instrument. A secondary purpose of the present study was to examine the presence of language biases across cultural backgrounds. A series of Differential Item Functioning tests were conducted to examine whether specific words were significantly more difficult for children of immigrants compared to Greeks. By applying the Rasch model, results indicated that most words were not significantly more challenging for immigrant students than for Greeks. Finally, a new ordering of the scale was proposed on the basis of calculated item difficulties.

Key words: Spelling, Normative data, Rasch model, Bias

1. *Address for correspondence:* Giorgos Sideridis, Department of Psychology, University of Crete, Rethymno, Greece 74100. e-mail: sideridis@psy.soc.uoc.gr

ΣΧΟΛΙΑ ΠΑΝΩ ΣΤΑ ΜΟΝΤΕΛΑ RASCH

Οι μέθοδοι για τον υπολογισμό των παραμέτρων σε μια μέτρηση κατά Rasch είναι αρκετοί σε αριθμό και μπορούν να μελετηθούν στη βιβλιογραφία σε κείμενα όπως αυτά των Wright και Masters (1982) καθώς και Linacre (1999, 2004). Προϋπόθεση για να είναι έγκυρη μια ανάλυση Rasch είναι τα ερωτήματα να προσαρμόζονται στην απλή λογική της, σύμφωνα με την οποία η επίδοση είναι το άθροισμα της ικανότητας των εξεταζομένων και (μείον) την δυσκολία των θεμάτων. Οι Bord και Fox (2015) συμβουλεύουν τους ερευνητές και τις ερευνήτριες να μην απορρίπτουν τα ερωτήματα-θέματα που δεν προσαρμόζονται σε αυτή την απλή λογική, αλλά να προσπαθούν να εξακριβώσουν τους θεωρητικούς λόγους πίσω από κάθε έλλειψη προσαρμογής.

Η ανάλυση παραγόντων και η μέτρηση κατά Rasch θεωρούνται -και είναι- ανόμοιες, λόγω των σχολών από τις οποίες προέρχονται. Η πρώτη, όπως είδαμε, προέρχεται από την κλασική θεωρία των μετρήσεων, ενώ η δεύτερη είναι παιδί της μοντέρνας θεωρίας των μετρήσεων. Η Ανάλυση Παραγόντων μοιάζει με την γραμμική παλινδρόμηση, με την διαφορά ότι στην πρώτη οι ερευνητές θα πρέπει να αποφασίσουν τον αριθμό, τις φορτίσεις αλλά, τη φύση ακόμα και τα ονόματα των κοινών παραγόντων. Για να το κάνουν αυτό εφαρμόζουν ορισμένους περιορισμούς κατά τη κατασκευή των μοντέλων. Θέτουν, για παράδειγμα, τον περιορισμό το μοντέλο να έχει έναν καθορισμένο αριθμό κοινών παραγόντων ή θέτουν κάποιους περιορισμούς σε σχέση ποιος κοινός παράγοντας, δηλαδή ποια λανθάνουσα μεταβλητή, σχετίζεται ισχυρότερα με ορισμένες μόνο από τις μεταβλητές και όχι με όλες. Στην ανάλυση κατά Rasch η γραμμική λογική είναι η ίδια και το μόνο που αλλάζει είναι η φύση των λανθανουσών μεταβλητών. Αυτά που προστίθεται εδώ είναι μόνο δύο στοιχεία, δηλαδή η δυσκολία του θέματος ή του ερωτήματος του τεστ και η ικανότητα του συμμετέχοντα ή της συμμετέχουσας.

Σε σύγκριση με την ανάλυση παραγόντων, η ανάλυση κατά Rasch έχει ορισμένα χαρακτηριστικά, τα οποία, σύμφωνα με τον Wright (1977), την κάνουν «ανώτερη». Πρώτον στην ανάλυση κατά Rasch η μέτρηση των θεμάτων-ερωτημάτων και των ατόμων γίνεται στην ίδια κλίμακα, σύμφωνα με τους Iramaneerat, Smith και Smith (2008). Δεύτερον αυτή η κλίμακα είναι ίσων διαστημάτων και όχι τακτικών τιμών όπως στις μετρήσεις της ανάλυσης παραγόντων. Τρίτον, η μέτρηση κατά Rasch έχει το χαρακτηριστικό της αμεταβλητότητας, όπως γράφει ο Engelhard (2013). Αμεταβλητότητα σημαίνει ότι η κατανομή της ικανότητας των ατόμων δεν εξαρτάται από την κατανομή δυσκολίας των ερωτημάτων-θεμάτων.

Στη βιβλιογραφία έχει συζητηθεί το θέμα του κατά πόσο η ανάλυση παραγόντων συνδέεται με την ανάλυση κατά Rasch (βλ. Holt, Duijn, & Boomsma, 2010; Takane & de Leeuw, 1987; Kamata & Bauer, 2008; Maydeu-Olivares, Cai, & Hernández, 2011; Wirth & Edwards, 2007). Υποστηρίζεται ότι στην περίπτωση που έχουμε μη γραμμική ανάλυση παραγόντων (π.χ. στην περίπτωση που έχουμε απαντήσεις

του τύπου «σωστό» - «λάθος», όπως στην έρευνα των Σιδερίδη και των συνεργατών του που είδαμε, οι παράμετροι της ανάλυσης παραγόντων μπορούν να μετατραπούν σε παραμέτρους ανάλυσης κατά Rasch (Kamata & Bauer, 2008). Ακόμα και στο ζήτημα της μέτρησης της αξιοπιστίας των μετρήσεων- στην ουσία μια διαδικασία που βρίσκεται στον πυρήνα της θεωρίας του πραγματικού σκορ- οι δύο μέθοδοι έχουν πολλά κοινά στοιχεία. Εκεί που οι δύο θεωρίες διαφέρουν πολύ είναι στην έννοια της μονοδιαστατικότητας, στην οποία θα σταθούμε.

Ένα κρίσιμο χαρακτηριστικό των τεστ, λοιπόν, τόσο στην ανάλυση παραγόντων (κλασική θεωρία των τεστ) όσο και στη μέτρηση κατά Rasch είναι, σύμφωνα με τον Thurstone (1931), η ύπαρξη μίας και μόνης διάστασης στα δεδομένα ή αλλιώς η «μονοδιαστατικότητα» (“unidimensionality” στην ξενόγλωσση βιβλιογραφία). Η ιδιότητα αυτή είναι απολύτως σημαντική κυρίως στη μέτρηση κατά Rasch, λόγω της αξίωσης της αμεταβλητότητας και της ειδικής αμεροληψίας των μετρήσεων (Brentari & Golia, 2007) για τις οποίες μιλήσαμε.

Στην ανάλυση παραγόντων κάθε απόφαση περί του βέλτιστου αριθμού των διαστάσεων είναι μέρος της διαδικασίας αυτής καθ’ αυτής και υπάρχουν πολλές μέθοδοι, καθεμιά με τον δικό της βαθμό πολυπλοκότητας, τις οποίες οι ερευνητές χρησιμοποιούν για να καταλήξουν στον βέλτιστο αριθμό παραγόντων (βλ. Hair, Anderson, Tatham, & Black, 1995). Στην ανάλυση κατά Rasch, όμως, η μονοδιαστατικότητα όχι μόνο είναι εκ των ουκ άνευ, αλλά επιπλέον οι ειδικοί δεν έχουν καταλήξει ακόμα σε ασφαλή συμπεράσματα για το πώς πρέπει αυτή να ελέγχεται. Κατ’ αρχάς, οι διαδικασίες ελέγχου της καλής εφαρμογής του μοντέλου μέσω των δεικτών Outfit και Infit, τις οποίες παρουσιάσαμε σε προηγούμενες ενότητες, δεν είναι κατάλληλες για την αξιολόγηση της μονοδιαστατικότητας. Αυτό συμβαίνει διότι σύμφωνα με τους Christensen, Engelhard και Salzberger (2012), όταν σε μια ανάλυση τα περισσότερα ερωτήματα μετράνε μία και μόνη διάσταση και κάποια ερωτήματα μετράνε μια άλλη διάσταση, τα δεύτερα επισημαίνονται εξ’ αρχής από τις μετρήσεις καλής εφαρμογής του μοντέλου Rasch ως ανόμοια και συνήθως αποσύρονται από την ανάλυση με ευθύνη των ίδιων των ερευνητών.

Ο Wright (στο Linacre, 2015) προτείνει στους ερευνητές να αξιολογούν την ύπαρξη της μονοδιαστατικότητας μέσω ανάλυσης κύριων συνιστωσών στα κατάλοιπα (residuals) του μοντέλου Rasch. Κατόπιν προτείνει να χωρίζεται το τεστ σε δύο μέρη, καθένα από εκ των οποίων να περιέχει μόνο τα ερωτήματα που εμφανίζουν υψηλή συνάφεια με καθεμιά από τις κύριες συνιστώσες (θετική ή αρνητική). Τέλος, προτείνει να υπολογίζεται ο συντελεστής συνάφειας μεταξύ των δύο μερών, στα οποία χωρίστηκε το αρχικό τεστ. Στην ιστοσελίδα του στατιστικού πακέτου Winsteps ο Linacre προτείνει μια τιμή Person r ίση με 0,57 ως όριο κάτω από το οποίο τα τα δύο μέρη του τεστ δεν μοιάζουν μεταξύ τους και ως συνέπεια ο ερευνητής συμπεραίνει έλλειψη μονοδιαστατικότητας.

Άλλοι τρόποι για να αξιολογηθεί η μονοδιαστατικότητα ενός τεστ στο πλαίσιο της ανάλυσης κατά Rasch είναι, σύμφωνα με την Verhelst (2001), το τεστ Martin-Löf (βλ. Gustafsson, 1980), το λεγόμενο Q2 τεστ του van der Wollenberg (1982), καθώς και η τεχνική του διαχωρισμού των ερωτημάτων (splitter-item-technique), επίσης του van den Wollenberg (ό. π.). Στο ίδιο πλαίσιο, οι Hambleton και Rovinelli (1986) έχουν συλλέξει και παρουσιάζουν περί τους 87 δείκτες για τον έλεγχο της μονοδιαστατικότητας γενικά και όχι αποκλειστικά στην ανάλυση κατά Rasch. Πάντως, η κύρια αντίληψη μέσα στην άτυπη επιστημονική κοινότητα των ψυχομετρών που χρησιμοποιούν σήμερα την ανάλυση κατά Rasch θα μπορούσε να συνοψιστεί σε αυτό που έχουν γράψει οι Reise, Cook, and Moore (2015, p. 13):

Το σημαντικό στην ανάλυση δεν είναι αν τα δεδομένα είναι μονοδιάστατα, αλλά ο βαθμός στον οποίο η υποτιθέμενη πολυδιαστατικότητα έρχεται σε σύγκρουση ή προκαλεί στρεβλώσεις στην εκτίμηση των παραμέτρων της ανάλυσης Rasch [δική μας μετάφραση].

Στο ίδιο πλαίσιο ο Everett Smith (2002, p. 206) θα γράψει ότι «αντί να ρωτάμε «μονοδιάστατο ή όχι» [για ένα τεστ], θα έπρεπε να ρωτάμε «από ποιο σημείο ενός συνεχούς και μετά αυτή η πολυδιαστατικότητα αποτελεί απειλή για τον υπολογισμό των παραμέτρων των ερωτημάτων και των ατόμων».

Αξίζει να αναφέρουμε ότι από τα τέλη της δεκαετίας του 1970 μέχρι σήμερα ο καθηγητής Harvey Goldstein, αρχικά από το London Institute of Education και τώρα από το Πανεπιστήμιο του Bristol, έχει αμφισβητήσει συνολικά την μονοδιαστατικότητα των μοντέλων Rasch, χρησιμοποιώντας επιχειρήματα από τον χώρο της θεωρητικής και εφαρμοσμένης Στατιστικής αλλά και από τους χώρους της Ψυχομετρίας και της Εκπαιδευτικής Αξιολόγησης (βλ. Goldstein & Blinkhorn, 1982· Goldstein, 1979a, 1979b). Τα επιχειρήματα αυτά δεν θα παρουσιαστούν εδώ, αλλά είναι σημαντικό ότι μέσω της αμφισβήτησης της μονοδιαστατικότητας και της τοπικής ανεξαρτησίας των ερωτημάτων των διεθνών ερευνών μαθητικής επίδοσης, ο Goldstein, όπως και άλλοι ερευνητές, έχει ασκήσει κριτική στο πρόγραμμα PISA (βλ. Goldstein, 2004· Kreiner & Christensen, 2014). Η σχετική συζήτηση συνεχίζεται σε κείμενα όπως αυτό του Goldstein (2015) και τις θέσεις των Panayides, Robinson και Tymms (2015). Ένα προτεινόμενο εισαγωγικό κείμενο στο θέμα της μονοδιαστατικότητας αποτελεί η εργασία του Wiberg (2012) στο περιοδικό *Educational Research and Evaluation*.

Οι Wright και Stone (1999) υποστηρίζουν ότι η ανάλυση αξιοπιστίας ενός τεστ μέσω για παράδειγμα του δείκτη Kuber και Richardson formula 20 ή του δείκτη alpha του Cronbach είναι πιο ακριβής όταν βρισκόμαστε σε περιβάλλον μέτρησης κατά Rasch παρά στο περιβάλλον της θεωρίας του πραγματικού σκορ. Αυτό γίνεται διότι τα τυπικά σφάλματα στην ανάλυση Rasch βασίζονται σε μετρήσεις οι οποίες είναι γραμμικοί –απλοί λογαριθμικοί– μετασχηματισμοί των αρχικών σκορ. Στην ανάλυση κατά Rasch το μέση διακύμανση του στατιστικού σφάλματος

βασίζεται στα τετράγωνα των υπολοίπων που βασίζονται με τη σειρά τους στον λογαριθμικό μετασχηματισμό των αρχικών τιμών κάθε ερωτήματος-θέματος του τεστ. Πιο συγκεκριμένα, στην ανάλυση κατά Rasch χρησιμοποιείται για τον υπολογισμό της αξιοπιστίας η μέση διακύμανση από όλα τα σφάλματα στο δείγμα αντί για την διακύμανση του σφάλματος του «μέσου ατόμου» (Wright & Stone, 1999).

Κλείνοντας την συζήτηση για την ανάλυση κατά Rasch, θα αναφερθούμε στους συντελεστές αξιοπιστίας, τόσο στην εν λόγω ανάλυση, όσο και στις αναλύσεις που βασίζονται στη θεωρία του πραγματικού σκορ. Όπως είναι γνωστό, οι συντελεστές αξιοπιστίας έχουν ως κατώτατο όριο το 0 και ως ανώτατο όριο το 1. Όμως, η μη γραμμική τους φύση προκαλεί δυσκολίες στην ερμηνεία τους. Για παράδειγμα, μια αύξηση του δείκτη αξιοπιστίας Cronbach α από 0,6 στο 0,7 δεν σημαίνει πρακτικά ίδια ποσοτικά βελτίωση όση θα σήμαινε μια αύξηση του ίδιου δείκτη από 0,7 στο 0,8, παρόλο που αριθμητικά η μεταβολή είναι ίδια. Για να αντιμετωπιστεί λοιπόν το ζήτημα αυτό που προφανώς προκύπτει την ύπαρξη ανώτατου ορίου «οροφής» στο 1, στην ανάλυση κατά Rasch χρησιμοποιούμε πολλές φορές τον λεγόμενο «δείκτη διαχωρισμού» τόσο για τους συμμετέχοντες όσο και για τα θέματα. Ο εν λόγω δείκτης είναι η τετραγωνική ρίζα ενός κλάσματος, στον αριθμητή του οποίου έχουμε τον συντελεστή αξιοπιστίας (τον δείκτη αξιοπιστίας Cronbach α που γίνεται Kuder - Richardson formula 20 στα διχοτομικά δεδομένα) και στον παρονομαστή το 1 πλην τον ίδιο συντελεστή αξιοπιστίας (βλ. Iramaneerat et al., 2008). Ο δείκτης διαχωρισμού των ατόμων (persons' separation index), ο οποίος συμβολίζεται με G_p , δείχνει πόσο διαφέρουν τα άτομα σε σχέση με την ικανότητά τους σε μονάδες τυπικού σφάλματος (Wright & Masters, 1982· Fisher, 1992, στο Iramaneerat et al., 2008). Δείχνει με άλλα λόγια πόσες διακριτές μετρήσεις ατομικών ικανοτήτων υπάρχουν στο δείγμα, όταν οι πολύ υψηλές και οι πολύ χαμηλές ικανότητες έχουν εξαιρεθεί. Τέλος, από τον δείκτη διαχωρισμού προκύπτει και ο δείκτης του αριθμού των διαφορετικών επιπέδων (strata) στις μετρήσεις τόσο των ατόμων, όσο και των ερωτημάτων. Ο δείκτης αυτός μοιάζει με τον δείκτη διαχωρισμού με τη διαφορά ότι συμπεριλαμβάνει τις πολύ υψηλές και τις πολύ χαμηλές βαθμολογίες (Linacre, 2013). Για παράδειγμα, ο αριθμός των διαφορετικών επιπέδων δυσκολίας των ερωτήσεων είναι το ένα τρίτο του τετραπλάσιου του δείκτη διαχωρισμού τους αυξημένου κατά 1). Ο δείκτης αυτός δείχνει πόσα επίπεδα δυσκολίας των θεμάτων μπορούμε να αναμένουμε στο δείγμα.

Βιβλιογραφικές αναφορές

- Allen, M.J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.
- Bollen, K. (2002). Latent variables in Psychology and the Social Sciences. *Annual Review of Psychology*, 53, 605–634. <http://doi.org/10.1146/annurev.psych.53.100901.135239>
- de Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. New York, NY: Guilford Press.
- Kellow, T. (2006). Using Principal Components Analysis in Program Evaluation: Some Practical Considerations. *Journal of MultiDisciplinary Evaluation*, 5, 89–107. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/viewFile/52/59
- Kolenikov, S., & Angeles, G. (2004). *The Use of Discrete Data in PCA : Theory , Simulations , and Applications to Socioeconomic Indices*. Moscow, Russia.
- Kreiner, S., & Christensen, K. B. (2013). Analyses of model fit and robustness. A new look at the PISA scaling model, underlying ranking of countries according to reading literacy. *Psychometrika*, 79(2), 210–231. <http://doi.org/10.1007/S11336-013-9347-Z>
- Linacre, J. M. (2005). Correlation coefficients: Describing relationships. *Rasch Measurement Transactions*, 19(3), 1028–1029. Retrieved from <http://www.rasch.org/rmt/>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Richmond, VA: Richmond, Byrd Press. Retrieved from <http://www.worldcat.org/title/estimation-of-latent-ability-using-a-response-pattern-of-graded-scores/oclc/193662>
- Sideridis, G. (2011). The effects of local item dependence on estimates of ability in the Rasch model. *Rasch Measurement Transactions*, 25(3), 1334–1336. Retrieved from <http://www.rasch.org/rmt/rmt253d.htm>
- Smith, L. (2002). *A tutorial on Principal Components Analysis*. Otago, New Zealand.
- Stoica, P., & Jansson, M. (2009). On maximum likelihood estimation in factor analysis-an algebraic derivation. *Signal Processing*, 89(6), 1260–1262. <http://doi.org/10.1016/j.sigpro.2009.01.002>
- Thomson, R. (1980). Maximum Likelihood Estimation of variance components. *Math. Operationsforsch*, 11, 545–561.
- Thurstone, L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433–451. Retrieved from http://www.brocku.ca/MeadProject/Thurstone/Thurstone_1925b.htm
- Κασσωτακης, Μ. (2013). *Η αξιολόγηση της επίδοσης των μαθητών: θεωρητικές προσεγγίσεις και πρακτικές*. Αθήνα: Γρηγόρη.
- Σιδερίδης, Γ., Μουζάκης, Α., Πρωτόπαπας, Α., & Σίμος, Π. (2008). Ψυχομετρική διερεύνηση μιας ορθογραφικής δοκιμασίας για μαθητές δημοτικού σχολείου. *Ψυχολογία*, 15(3), 290–315. Retrieved from http://www.ilsp.gr/homepages/protopapas/pdf/Sideridis_etal_2008_Psychologia.pdf